Sec. B

Journal of the AMERICAN STATISTICAL ASSOCIATION

UNIVERSITY OF MICHIGAN
DEC 28 1951
BUSINESS ADMINISTRATION LIBRARY

DECEMBER 1951

Redistribution of Population, 1940 to 1950 ..... *Henry S. Shryock, Jr.* 417

Changes in Income Distribution in the United States ..... *Herman P. Miller* 438

Results of Alternative Statistical Treatments of Sample Survey Data ..... *L. R. Klein and J. N. Morgan* 442

An Economic Appraisal of Price Measures ..... *Richard B. Heflebower* 461

Mathematical Models in Psychological Scaling ..... *C. H. Coombs* 480

Yates' Correction and the Statisticians ..... *Frank Adler* 490

The Distribution of the Quotient of Ranges in Samples from a Rectangular Population ..... *Paul R. Rider* 502

The Measurement and Prediction of Attitudes ..... *William G. Madow* 508

Erratum: On the Determination of Sample Sizes in Designing Experiments ..... *A. M. Mood* 515

Reprints of Abstracts in Statistical Methodology 516

BOOK REVIEWS by F. Stuart Chapin, Jerome Cornfield, David Durand, Bruno de Finetti, Harry P. Hartkemeier, Gopinath Kallianpur, Ernest Kurnow, Margaret Merrell, Loring G. Mitten, Lincoln E. Moses, C. Ashley Wright ... 518

PUBLICATIONS RECEIVED 538

Index to Volume 46, 1951 539

VOLUME 46 NUMBER 256

American Statistical Association

Organized November 27, 1839, Incorporated 1841

The American Statistical Association is a scientific and educational organization. Its membership is not confined to professional statisticians but includes economists, business executives, research directors, government officials, university professors, and others who are seriously interested in the application of statistical methods to practical problems, in the development of more useful methods, and in the improvement of basic statistical data. Engineers, mathematicians, biologists, actuaries, sociologists, psychologists, and representatives of many other professions are included in the membership of the Association.

Annual Dues

Regular membership	$ 8*	Member's subscription to Biometrics	4
Student membership	4	Contributing membership	25
Introductory membership (first dues payment of members under 30 years old)	4	Institutional membership, minimum	100

* Includes $5 for a year's subscription to the *Journal of the American Statistical Association*, and $1.00 for a year's subscription to *The American Statistician*.

Subscription rate, $8 per year. Prices for back issues available on request. A cumulative Index to Volumes 1–34, 1888–1939, and Annual Indexes thereafter may be obtained from the Secretary.

For further information about the Association and membership application forms, write the Secretary, 1108 16th Street, N. W., Washington 6, D.C.

The present Institutional Members are:

American Telephone and Telegraph Company
Arabian American Oil Company
Armstrong Cork Company
Chrysler Corporation
Cities Service Company
Deere and Company
Dun and Bradstreet, Inc.
Equitable Life Assurance Society of U.S.
Ford Motor Company
General Motors Corporation
Industrial Surveys Company
International Business Machines, Inc.
The McBee Company
Metropolitan Life Insurance Co.
The National Bank of Detroit
Standard Oil Company (Indiana)
Standard Oil Company (New Jersey)
Sun Oil Company
Western Electric Company, Inc.

Published Quarterly by the AMERICAN STATISTICAL ASSOCIATION

Publication Office: 450 Ahnaip Street, Menasha, Wisconsin. Editorial Office: 307 Haskell Hall, University of Chicago, Chicago 37, Illinois. Business Office: 1108 16th St. N.W., Washington 6, D.C. Acceptance for mailing at special rate of postage provided for in the Act of February 28, 1925, embodied in paragraph 4, section 538, P. L. & R., authorized March 23, 1926. Entered as second class matter at the post office at Menasha, Wisconsin.

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

VOLUME 46 DECEMBER 1951 NUMBER 256

ARTICLES

Redistribution of Population, 1940 to 1950 . HENRY S. SHRYOCK, JR. 417
Changes in Income Distribution in the United States . . . . . HERMAN P. MILLER 438
Results of Alternative Statistical Treatments of Sample Survey Data . . . . . L. R. KLEIN AND J. N. MORGAN 442
An Economic Appraisal of Price Measures . RICHARD B. HEFLEBOWER 461
Mathematical Models in Psychological Scaling . . . C. H. COOMBS 480
Yates' Correction and the Statisticians . . . . . FRANZ ADLER 490
The Distribution of the Quotient of Ranges in Samples from a Rectangular Population . . . . . PAUL R. RIDER 502
The Measurement and Prediction of Attitudes . WILLIAM G. MADOW 508
Erratum: On the Determination of Sample Sizes in Designing Experiments . . . . . A. M. MOOD 515
Reprints of Abstracts in Statistical Methodology 516
Index of *Journal*, Volume 46, 1951 (Nos. 253–256) 539
Articles, by Author 545
Book Reviews, by Author 546

BOOK REVIEWS

STOUFFER, SAMUEL A., LOUIS GUTTMAN, EDWARD A. SUCHMAN, PAUL F. LAZARSFELD, SHIRLEY A. STAR, AND JOHN A. CLAUSEN, *Measurement and Prediction* . . . . . WILLIAM G. MADOW 518
MITCHELL, WESLEY C., *What Happens During Business Cycles* . . . . . C. ASHLEY WRIGHT 518
MACNIECE, E. H., *Production Forecasting, Planning, and Control* . . . . . LORING G. MITTEN 521
ZIMMERMANN, ERICH W., *World Resources and Industries* 522
UNITED NATIONS, *Statistical Yearbook, 1949–50, Second Issue* 523
ROSANDER, A. C., *Elementary Principles of Statistics*. . . . . HARRY P. HARTKEMEIER 523
NEISWANGER, WILLIAM A., FLOYD B. HAWORTH, AND WILLIAM L. LEAVITT, *Laboratory Manual for Elementary Statistical Methods as Applied to Business* . . . . . ERNEST KURNOW 524
TRELOAR, ALAN E., *Biometric Analysis: An Introduction* . . . . . JEROME CORNFIELD 525
LINDQUIST, E. F., ed., *Educational Measurement* . . LINCOLN E. MOSES 527
COHN, ALFRED E. AND CLAIRE LINGG, *The Burden of Diseases in the United States* . . . . . MARGARET MERRELL 528
OHLIN, LLOYD E., *Selection for Parole* . . . . . DAVID DURAND 529
WHITE, VIRGINIA KANN, *Measuring Social Need: The Method of Standard Scores Applied to the Measurement of Need in Cuyahoga County* . . . . . F. STUART CHAPIN 530
CARNAP, RUDOLPH, *Logical Foundations of Probability* . . . . . BRUNO DE FINETTI 532
CARNAP, RUDOLPH, *The Nature and Application of Inductive Logic*. . . 534
MUNROE, M. E., *Theory of Probability*. . . GOPINATH KALLIANPUR 535
Trabajos de Estadistica 536
Revue Statistique 536
Revue Economique. 537
PUBLICATIONS RECEIVED. 538

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

The Editors welcome the submission of manuscripts for possible publication. They should be typewritten entirely double-spaced, including footnotes, and two copies should be sent to the Editor, W. Allen Wallis, 207 Haskell Hall, University of Chicago, Chicago 37. Books for review should be sent to the same address. Unsolicited book reviews are not accepted, but suggestions of titles for review are welcome.

EDITOR

W. ALLEN WALLIS, *University of Chicago*
ASSISTANT TO THE EDITOR: MARGARET A. LABADIE

ASSOCIATE EDITORS

ALBERT H. BOWKER
Stanford University
SOLOMON FABRICANT
New York University
HOWARD L. JONES
Illinois Bell Telephone Co.
LESTER S. KELLOGG
Deere and Company
WILLIAM G. MADOW
University of Illinois
FREDERICK MOSTELLER
Harvard University

ADVISORY PANEL OF FORMER EDITORS

WILLIAM G. COCHRAN (1945–50)
Johns Hopkins University
WILLIAM F. OGBURN (1920–1925)
University of Chicago
FRANK A. ROSS (1926–34, 41–45)
Thetford, Vermont
FREDERICK F. STEPHAN (1935–40)
Princeton University

EDITORIAL COLLABORATORS

ARMEN A. ALCHIAN
University of California (Los Angeles)
R. L. ANDERSON
North Carolina State College
KENNETH J. ARROW
Stanford University
GERTRUDE BANCROFT
Bureau of the Census
ERNEST T. BAUGHMAN
Federal Reserve Bank of Chicago
COLIN R. BLYTH
University of Illinois
DONALD J. BOGUE
Miami University
MARTIN A. BRUMBAUGH
Bristol Laboratories, Inc.
CHARLES F. CANNELL
University of Michigan
UTTAM CHAND
Boston University
DOUGLAS G. CHAPMAN
University of Washington
W. W. CHARTERS, JR.
University of Illinois
MORRIS A. COPELAND
Cornell University
JEROME CORNFIELD
National Cancer Institute
ANDREW T. COURT
General Motors Corporation
JOSEPH S. DAVIS
Stanford University
ROBERT DORFMAN
University of California (Berkeley)
HAROLD F. DORN
National Institute of Health
PAUL S. DWYER
University of Michigan
WALTER EBHLING
Federal Crop Reporting Service
CHURCHILL EISENHART
National Bureau of Standards
HOPE T. ELDRIDGE
United Nations
WILLIAM FELLER
Princeton University
A. L. FINKNER
North Carolina State College
WALTER FIREY
University of Texas
LESTER R. FRANKEL
Alfred Politz Research, Inc.
HAROLD A. FREEMAN
Massachusetts Institute of Technology
MILTON FRIEDMAN
University of Chicago
WILFRED GIBSON
University of Illinois
HENRY GOODE
Stanford University
ROE GOODMAN
University of Michigan
BERT L. GREEN
Massachusetts Institute of Technology
THOMAS N. E. GREVILLE
Public Health Service
MARGARET J. HAGOOD
University of Wisconsin
K. W. HALBERT
American Telephone and Telegraph Company
FRANK HANNA
Duke University
MORRIS H. HANSEN
Bureau of the Census
HARRY P. HARTKEMEIER
University of Missouri
CECIL HASTINGS
The Rand Corporation
PHILIP M. HAUSER
University of Chicago
RICHARD D. HEFLEBOWER
Northwestern University
ROGER W. HEYNS
University of Michigan
J. L. HODGES, JR.
University of Chicago
EARL E. HOUSEMAN
Bureau of Agricultural Economics
HARRY M. HUGHES
University of California (Berkeley)
LEONID HURWICZ
University of Illinois
EMIL H. JEBE
Iowa State College
R. J. JESSEN
Iowa State College
LYLE JOHNSON
Headquarters, U. S. Air Force
WILLIAM O. JONES
Stanford University
BERNARD KARPINOS
Office of the Surgeon General
OSCAR KEMPTHORNE
Iowa State College
MAURICE G. KENDALL
University of London

LAWRENCE KLEIN
National Bureau of Economic Research
LLOYD A. KNOWLER
University of Iowa
ROBERT KOSELKA
Harvard University
GEORGE M. KUZNETS
University of California (Berkeley)
PAUL LAZARSFELD
Columbia University
H. GREGG LEWIS
University of Chicago
GERALD LIEBERMAN
Stanford University
CLARENCE LONG
Johns Hopkins University
EUGENE LUKACS
National Bureau of Standards
ELI S. MARKS
Bureau of the Census
PHILIP J. MCCARTHY
Cornell University
MARGARET MERRELL
Johns Hopkins University
K. E. MILLER
Armour and Company
ALEXANDER M. MOOD
The Rand Corporation
GEOFFREY H. MOORE
National Bureau of Economic Research
LINCOLN MOSES
Columbia University
BRUCE D. MUDGETT
University of Minnesota
B. B. MURDOCH
Prudential Insurance Company of America
R. B. MURPHY
Carnegie Institute of Technology
WILLIAM A. NEISWANGER
University of Illinois
HAROLD NISSELSON
Bureau of the Census
HORACE W. NORTON
University of Illinois
ALFRED OXENFELDT
Hofstra College
EDWARD PAULSON
University of Washington
PAUL R. RIDER
Washington University
HARRY V. ROBERTS
University of Chicago
A. C. ROSANDER
Bureau of Internal Revenue
BARKEV S. SANDERS
Social Security Board
OSCAR N. SERBEIN, JR.
Columbia University
HENRY S. SHRYOCK, JR.
Bureau of the Census
IRVING H. SIEGEL
Twentieth Century Fund
H. FAIRFIELD SMITH
North Carolina State College
JOHN H. SMITH
American University
HERBERT SOLOMON
Office of Naval Research
ROBERT SOLOW
Massachusetts Institute of Technology
SHIRLEY A. STAR
National Opinion Research Center
JOSEPH STEINBERG
Bureau of the Census
FREDERICK F. STEPHAN
Princeton University
S. S. STEVENS
Harvard University
FRED L. STRODTBECK
Yale University
LORIE TARSHIS
Stanford University
BENJAMIN J. TEPPING
Bureau of the Census
LEDYARD TUCKER
Educational Testing Service
JOHN W. TUKEY
Princeton University
MELVILLE J. ULMER
American University
WILLIAM S. VICKREY
Columbia University
HELEN M. WALKER
Columbia University
PAUL P. WALLRABENSTEIN
Bureau of Agricultural Economics
ALFRED N. WATSON
Curtis Publishing Company
MAX WOODBURY
Princeton University
JANE WORCESTER
Harvard School of Public Health
ELMER J. WORKING
University of Illinois

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION

Number 256 DECEMBER 1951 *Volume 46*

REDISTRIBUTION OF POPULATION: 1940 TO 1950*

HENRY S. SHRYOCK, JR.
U. S. Bureau of the Census

THE forced movements of masses of European and Asian people during and after the last World War challenge our imaginations, but the less spectacular movements inside the United States have been of similar magnitude and have probably resulted in rather more net change in geographic distribution. In contrast to the flight of refugees, the forced transfers of ethnic groups, and the deportation of minorities, American internal migration has consisted of voluntary moves by individuals or families, motivated largely by economic considerations. In the absence of violence, it has also been possible to measure the shifts of our own population considerably more accurately. In this same period, all the components of population change were quite different from what they had been in the depressed 'thirties. All differed relative to the initial population in the various geographic areas. As usual, however, the differing patterns of migration rates and rates of natural-increase tended to work against each other as areas of high fertility and low economic opportunity sent out to areas of low fertility and high economic opportunity many more migrants than they received.

INEQUALITIES OF POPULATION CHANGE

The 19 million increase in the population of the United States between 1940 and 1950 had not been equalled in any previous decade, and yet many areas experienced actual decreases in population. Among the States, the percentage change ranged from +53.3 for California to −4.4 for Oklahoma. Among the more notable effects of this redistribution of population among the States will be the shifts in the House of Representatives. Fourteen seats will be shifted, seven States gaining and nine States losing seats. The geographic center of the House is shifting westward.

* Paper read at the 110th Annual Meeting of the American Statistical Association, at Chicago, Ill., December 27, 1950. The writer is indebted to Donald S. Akers and to Calvin L. Beale of the Bureau of the Census for their assistance, particularly in the compilation of some of the statistics.

Distribution of counties by intercensal change[1]

All counties.—Among the 3103 counties or equivalent areas, almost half lost population [1]. The distribution by percentage change is given in Table 1. It will be noted that in each geographic division there was a wide range of change. Even in the Pacific division, which increased by 49 per cent as a whole, about one-tenth of the counties lost population; and even in the West North Central division, which increased by only 4 per cent as a whole, about one-twentieth gained by 20 per cent or more. The rate of change will not be compared here among the divisions or among the States, but later on a better unit for geographic analysis will be used.

TABLE 1

DISTRIBUTION OF COUNTIES ACCORDING TO PERCENTAGE CHANGE IN POPULATION, BY GEOGRAPHIC DIVISION: 1940 TO 1950

(1950 figures are preliminary)

Division	Total counties	Per cent of total counties							
		Loss				Gain			
		20 and over	10 but under 20	5 but under 10	0 but under 5	0 but under 5	5 but under 10	10 but under 20	20 and over
United States	3,103	6.1	17.0	13.0	13.5	12.6	9.9	11.8	16.1
New England	67	—	1.5	3.0	9.0	26.9	25.4	25.4	9.0
Middle Atlantic	150	—	3.3	5.3	12.0	16.0	27.3	22.7	13.3
East North Central	436	0.9	7.6	10.8	14.2	17.4	15.4	19.3	14.4
West North Central	620	5.2	27.6	20.0	19.2	10.2	6.0	6.6	5.3
South Atlantic	582	1.9	11.2	11.9	13.7	15.1	11.3	13.7	21.1
East South Central	364	4.4	24.2	17.9	16.2	14.6	6.6	8.0	8.2
West South Central	470	21.3	23.8	10.9	7.4	5.7	6.0	7.7	17.2
Mountain	281	9.3	17.8	11.7	13.2	10.7	6.4	10.0	21.0
Pacific	133	1.5	3.0	3.0	3.0	7.5	6.8	12.8	62.4

Some of the counties with high rates of growth were considerably affected by the new method of enumerating college students. Prior to 1950 students were counted at their parental homes unless they were employed in the college community, were married, or said they had no home except in the college community. In 1950, all college students were supposed to have been enumerated where they were living while attending college. Many students, of course, live in their parental homes so were unaffected by the change. Even under the old procedure, college

[1] All figures other than those for states are preliminary and exclude 841,769 persons not yet allocated within the states at the time of this writing.

communities would have had fairly large increases because of the influx of ex-servicemen, who were more likely to be married and independent financially, and because of the expansion of staff and of service industries resulting from increased enrollment. This change in enumerative procedure had most effect upon the percentage change of otherwise relatively small counties that contained a large college or university. For example, Tompkins County, N. Y., the seat of Cornell University, increased by 39.8 per cent; and Centre County, Pennsylvania, where Pennsylvania State College is located, increased by 24.8 per cent. It will never be possible to determine exactly what part of the intercensal changes are due to this factor, but perhaps estimates can be made later.

Metropolitan and nonmetropolitan counties

More than half of the population of the United States was living in one of the 168 standard metropolitan areas in 1950 [2]. (So that a classification could be made including all the counties, the standard metropolitan areas in New England, which are defined on a town basis, are approximated here on a county basis; as was also done when the State economic areas were delimited.) While the population in the counties of the standard metropolitan areas was increasing by more than twenty per cent, that in the other counties was increasing by only about six per cent. The concentration of population in metropolitan areas, which slowed down during the depression, was greatly accelerated by the war and the postwar period of prosperity.

Following a technique of analysis recently used by Bogue [3], we may classify the metropolitan counties by the size of the central city in 1940 and the nonmetropolitan counties by the size of the largest urban place in 1940. The results are presented in Table 2. Every size category of metropolitan counties had a higher rate of increase than the national average for all counties. The metropolitan counties with central cities of intermediate size experienced somewhat more population growth on the average than did those with either very large or small central cities. Among nonmetropolitan counties, the rate of growth was greater the larger the most populous place in the county. The combined counties without an urban place in 1940 lost population. Only those nonmetropolitan counties with an urban place of at least 25,000 in 1940 gained population at above the national average rate. This class compared favorably with the metropolitan counties. Many new standard metropolitan areas were recruited from this class, of course.

TABLE 2

CHANGE IN THE POPULATION OF METROPOLITAN AND NONMETROPOLITAN COUNTIES OF THE UNITED STATES CLASSIFIED BY SIZE OF LARGEST INCORPORATED PLACE: 1940–1950

Type of county by size of largest place	1950* population	1940 population	Amount of change	Percentage change
Total	149,855,591	131,669,275	18,186,316	13.8
Metropolitan Counties† by size of central city of the Metropolitan Area in 1940:	82,514,342	68,425,343	14,088,000	20.6
1,000,000 and over	29,280,369	24,979,735	4,300,634	17.2
500,000 to 999,999	15,350,352	12,856,068	2,494,284	19.4
250,000 to 499,999	12,272,757	9,700,703	2,572,054	26.5
100,000 to 249,999	14,571,901	11,734,149	2,837,752	24.2
50,000 to 99,999	11,038,963	9,154,688	1,884,275	20.6
Nonmetropolitan Counties by size of largest urban place in 1940:	67,341,249	63,243,932	4,097,317	6.5
25,000 to 49,999	10,307,604	8,482,738	1,824,866	21.5
10,000 to 24,999	16,879,497	14,894,745	1,984,752	13.3
5,000 to 9,999	14,342,943	13,574,508	768,435	5.7
2,500 to 4,999	12,345,053	12,251,584	93,469	0.8
No urban place	13,466,152	14,040,357	−574,205	−4.1

* Preliminary counts of the population of counties as of April 1, 1950 from U. S. Bureau of the Census, Series PC-3, No. 4. Approximately 700,000 individual census reports are excluded from these figures.

† Counties classed as metropolitan in Series PC-3, No. 3 except that counties in metropolitan areas the central city of which was less than 50,000 in 1940 are classed as nonmetropolitan. New England counties classed as metropolitan or nonmetropolitan according to whether in metropolitan or nonmetropolitan State Economic Areas.

State Economic Areas

Figure 1 is a map showing the percentage change from 1940 to 1950 for the 443 state economic areas [4]. State economic areas are groups of counties that are similar with respect to a large number of demographic, social, agricultural, and industrial indices [5]. These areas reveal patterns that are concealed in areas as large as states and are difficult to detect among the mass of detail for the 3103 counties.

Metropolitan Economic Areas. First, we may note on the map the small scattered areas of rapid growth that represent metropolitan state economic areas.[2] Almost all these metropolitan areas increased in popu-

[2] Metropolitan state economic areas are the standard metropolitan areas that had a central city of 50,000 or more and a total population of 100,000 or more, both *prior* to 1950. Where such a standard metropolitan area is in two or more states, each part is a separate state economic area. The standard metropolitan areas in New England, which are made up of towns and cities, are approximated on a county basis to constitute state economic areas.

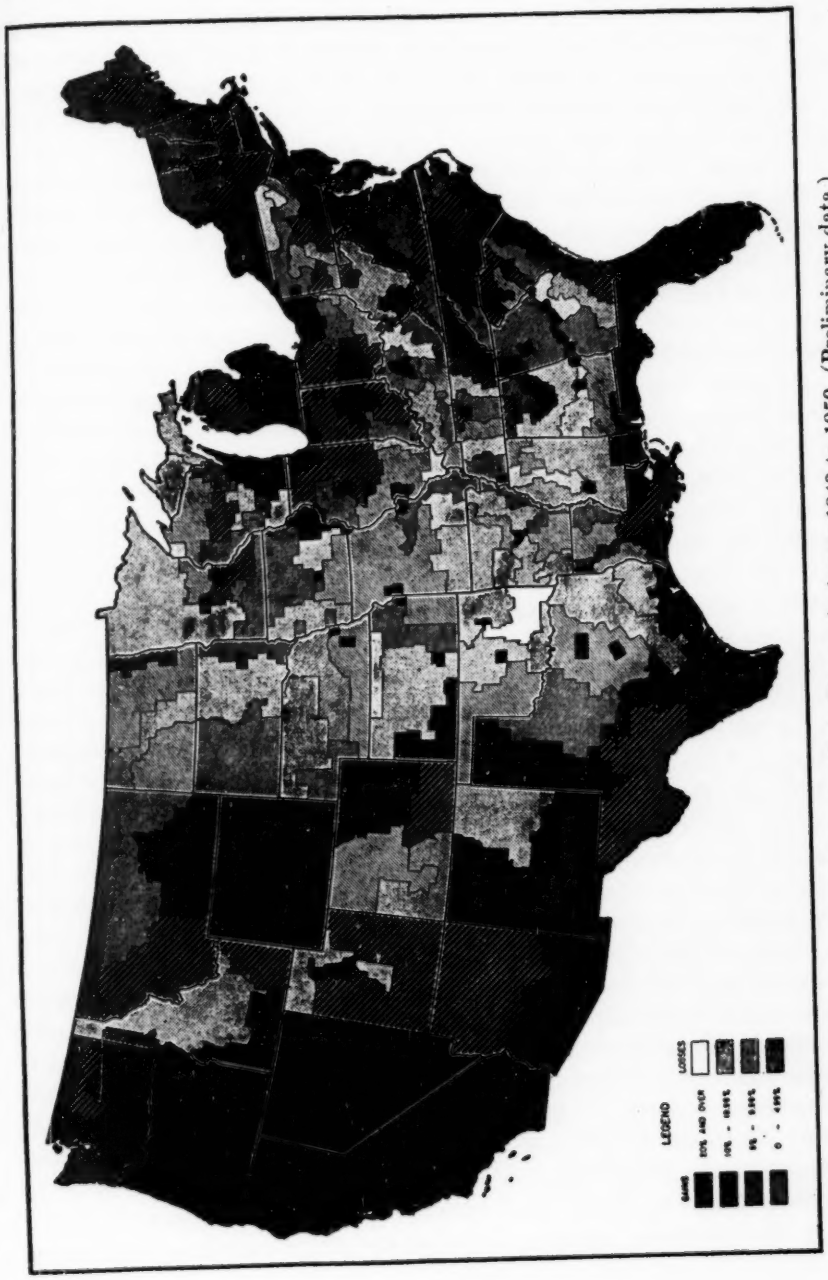


FIGURE 1. Per cent change by State Economic Areas: 1940 to 1950. (Preliminary data.)

lation, and many increased by one-third or more. In what follows, gains or losses of 20 per cent or more will be described as "heavy," those of 10 to 19 per cent as "moderate," those of 5 to 9 per cent as "slight," and populations changing by less than 5 per cent will be described as "static."

Along the Atlantic Coast, the rate of increase tends to be greater as we proceed from north to south. Above Delaware Bay, only the Stamford-Norwalk area experienced a heavy increase; and the population of the Fall River-New Bedford area was relatively static. From Wilmington and Baltimore south, all the increases were at least 20 per cent. The Miami, Norfolk-Portsmouth, and Jacksonville areas increased by about 80, 60, and 40 per cent, respectively. All of the five metropolitan state economic areas on the Gulf Coast also had heavy increases, ranging from 23 per cent for New Orleans to 61 per cent for Mobile. On the Pacific Coast, there were even heavier gains, ranging from Portland's 40 per cent to San Diego's 85 per cent. The picture along the Great Lakes is less straight-forward. Erie, Detroit, and Muskegon had heavy gains, but most of the other metropolitan areas had moderate gains; and the Duluth-Superior area had slight change. Some of the port areas reached population peaks during the war and afterwards dropped off somewhat as naval and shipbuilding installations were closed.

Looking at the inland metropolitan areas, we find relatively fewer heavy increases. In some broad regions, however, all of the inland metropolitan areas had heavy gains; this trend is true of the Southwest, the Rocky Mountains, and the Central Valley and other inland parts of California. Some of the most rapid growth in the nation occurred here, for example, 77 per cent in the Phoenix area and 74 per cent in the San Bernardino area. In the Southeastern Piedmont, six out of seven metropolitan areas had heavy gains (although the largest was only 35 per cent); and the seventh, Winston-Salem, had a moderate growth.

In the Appalachians as along the Atlantic Coast, the growth of metropolitan areas tended to be greater in the South. The Scranton and Wilkes-Barre–Hazleton areas, in the depressed anthracite region, were the only metropolitan state economic areas in the nation with losses of more than 10 per cent. Furthermore, the Altoona, Johnstown, and Wheeling areas were static, and the Pittsburgh area had only a slight gain. Moderate to heavy gains occurred south of the Mason-Dixon Line, with the Knoxville area (which contains Oak Ridge) increasing by more than one-third.

In the Northeast between the coast and the mountains, there were

slight to moderate increases. Inland from the Great Lakes, Midwestern areas were fairly evenly divided between those with moderate and those with heavy increases, with only one below this range. Both the largest and the smallest increases were to be found near the western limit of the Corn Belt—Wichita, which gained by more than half, and Sioux City, which was static. Nearby Wright Field contributed to Dayton's growth of over one-third. Finally, from the Ohio River down into the Deep South but outside the mountains and coastal plain, there were moderate to heavy gains except for the Huntingdon-Ashland area, which had only a slight gain. The Columbus, Ga., and Memphis metropolitan areas increased by about one-third.

In summary, there tended to be heavier growth among the metropolitan areas along the coast and around the Great Lakes than among the inland metropolitan areas. The younger metropolitan areas of the South and West tended to grow more than the maturer areas of the Northeast and Middle West. This latter generalization would be reinforced if we included the areas that first qualified as metropolitan in 1950. Most of these were areas of heavy increase in the South and West.

Nonmetropolitan Economic Areas. A tendency toward more rapid growth along the Atlantic, Pacific, and Gulf Coasts than in the heart of the country was characteristic of nonmetropolitan as well as of metropolitan state economic areas. There were heavy gains in the hinterland coastal areas of California, Oregon, and Texas; Kitsap County, Washington; the delta area at the mouth of the Mississippi River; the Gulf Coast of Mississippi; most of Florida; North Carolina south of Cape Hatteras; the lower Virginia "Peninsula"; southern Maryland; New Jersey; and as far north as Naragansett Bay in Rhode Island. There are important military installations in some of these areas. Areas of moderate increase comprise the northeastern shore of Puget Sound, southwestern Louisiana, the Alabama Gulf Coast, a thinly-settled belt across northern Florida from Gulf to Ocean, the South Carolina Low Country, the Virginia "South Side," southern Delaware, eastern Connecticut, and the coast of New Hampshire. Gains were slight in northwestern Washington, south central Louisiana, Tidewater Virginia from the Potomac to the York, the Eastern Shore of Maryland, and southern Maine. Finally, the North Carolina Tidewater above Hatteras, the Eastern Shore of Virginia, and northeastern Maine were static areas. All the areas in these last two growth classes were highly rural.

Around the Great Lakes, the heavy gains occurred in the industrial area of north central Ohio, in the counties bordering the Detroit metropolitan area, and on Lake Michigan between Muskegon and Gary. Gains were slight to moderate from Green Bay down to the Chicago area, in the other parts of southern Michigan that front on Lake Michigan or Lake Huron, and eastward to include most of the shore of Lake Ontario. The Great Lakes economic areas to the northward were static as to population change except for part of the Upper Peninsula of Michigan that had a moderate loss. Thus nonmetropolitan population increase around the Great Lakes tended to be greatest near the Chicago, Detroit, and Cleveland metropolitan areas and to fall off progressively to the east, west, and north.

The economic areas of Michigan, Wisconsin, and Minnesota around Lake Superior are part of the "Lake States Cut-over Area," which in 1935 was characterized as one of "six rural problem areas" by the Federal Emergency Relief Administration [6]. The state economic areas further inland in the Cut-over also had very little population change.

When we follow the Appalachians through the pertinent nonmetropolitan economic areas down from New York through Alabama, we find none with heavy gains and none with moderate or heavy losses. Two mountainous areas in Pennsylvania, one characterized by anthracite and the other by bituminous coal mining, had slight population losses. Furthermore, a great many areas were static; and, since they were generally areas of high fertility, they must have had substantial outmigration during the decade. The moderate gains begin with the rich Blue Ridge farming area of south central Pennsylvania and include the southwestern coal fields of West Virginia, the Shenandoah Valley, the North Carolina Blue Ridge, the easternmost state economic areas of Tennessee, and small parts of Georgia and Alabama.

Of the predominantly dairying areas that are neither coastal nor mountainous, most were static or had slight gains. A few such areas had moderate gains. One in Illinois had a heavy gain, and one in Minnesota had a moderate loss. These dairying economic areas ranged geographically from Maine to Minnesota and south to the Capital's milkshed in Virginia.

The Corn Belt, which extends from Ohio northwesterly into South Dakota, contained mostly state economic areas with slight losses to slight gains in population. A few had moderate gains, and one area in southern Michigan experienced a heavy increase. The only Corn Belt area with a loss of more than 10 per cent was in northern Kansas.

The major wheat-producing state economic areas are in the Great Plains and in the eastern Columbia River basin. Population trends here were quite variable, ranging from slight losses to heavy gains. Eight of the 19 areas of this type were static. The heavy gains occurred in Washington and Oregon and in the Texas Panhandle. (The growth of the natural gas industry was a factor in the last of these.)

The range livestock region contains state economic areas in both the Great Plains and the Rocky Mountains. Those in the plains ranged from moderate losses to moderate gains, those in the mountains from a slight loss to heavy gains. The heavy gains occurred in New Mexico, Arizona, and Nevada.

Turning east again, we may examine the predominantly tobacco-growing state economic areas. Of the 11 of these, only one is coastal and another mountainous. The heavy gain in southern Maryland is primarily attributable to military installations. Areas in the Piedmont, sand hills, and valleys ranged from static to those with moderate population growth.

The vast Cotton Belt extends from North Carolina to Texas and includes about 60 of the state economic areas. Population in about one-third of these was static. About one-third each had from slight to heavy losses, and from slight to heavy gains. Many of the areas of moderate loss were in the Black Belt, in the heart of the cotton empire.

About half of the general farming areas are in the mountains, and most of the remainder are inland. The few extant areas of self-sufficing agriculture are of this type. The mountainous general farming economic areas were mostly static or had slight population gains. The nonmountainous inland areas were mostly static or had slight population losses.

The state economic areas where fruit and truck crops predominate are mostly coastal. All but one experienced a gain of at least five per cent, and many had heavy gains. The coastal areas of Louisiana and Texas that specialize in rice growing had population gains ranging from slight to heavy.

It may be noted also that the former Dust Bowl has recovered its population losses of the 'thirties. The state economic areas that approximate the areas that had had major wind erosion [6] had a total population of 4,386,000 in 1930, 4,238,000 in 1940, and 4,398,000 in 1950—a decrease of 3.4 per cent from 1930 to 1940, and an increase of 3.8 per cent from 1940 to 1950. Better weather and a sustained demand have brought back prosperity to the wheat farmers and stockmen of

these areas, particularly in southwestern Kansas. Also new oil and natural gas fields have been developed in the Texas Panhandle.

This description of population change among nonmetropolitan state economic areas has been made mainly by classifying them as to type of farming and as to topography. A fairly wide range of change was found within each type. Presumably this range would be considerably narrowed if we could cross-classify the state economic areas further by urban-rural residence. Data from the 1950 Census for making this classification were not available at the time of writing.

Suburbanization

The data so far examined have indicated a general trend of concentration of population in metropolitan areas and along the coasts. This process has been going on for some decades, but the prosperous 'forties saw a great acceleration as compared with the depressed 'thirties. Within metropolitan areas, the well-known process of decentralization—from central city to suburbs—has also continued, however.

Let us return to the 168 standard metropolitan areas to examine this process of suburbanization. The preliminary figures show that while the central cities as a group were increasing by 13.0 per cent, the outlying parts of the metropolitan areas were increasing by 34.7 per cent [2]. Nearly half of the population increase of the nation took place in these suburban areas. In only 37 of 168 areas did the outlying parts grow less rapidly than the central city. In some of these exceptions, the central city had annexed part of its suburbs during the decade.

It is true that these relationships are affected by such arbitrary factors as the area of counties and the practices of state legislatures with respect to size of areas to be incorporated as municipalities and with respect to annexations. Nonetheless, the data indicate the extent to which the inhabitants of the economic city have their homes outside the boundaries of the political city and are thus in a large degree independent of its administration. Studies now being made at the Scripps Foundation and elsewhere are showing that, in the period 1935 to 1940, the in-migrants to the suburbs came largely from the central city but in-migrants to the central city came largely from outside the metropolitan area [7].

THE ROLES OF NATURAL INCREASE AND MIGRATION

What were the roles of natural increase and migration in effecting the redistribution of population during the 'forties? Figures 2 and 3 show

FIGURE 2. Natural increase, 1940 to 1950, as per cent of 1940 population, by States

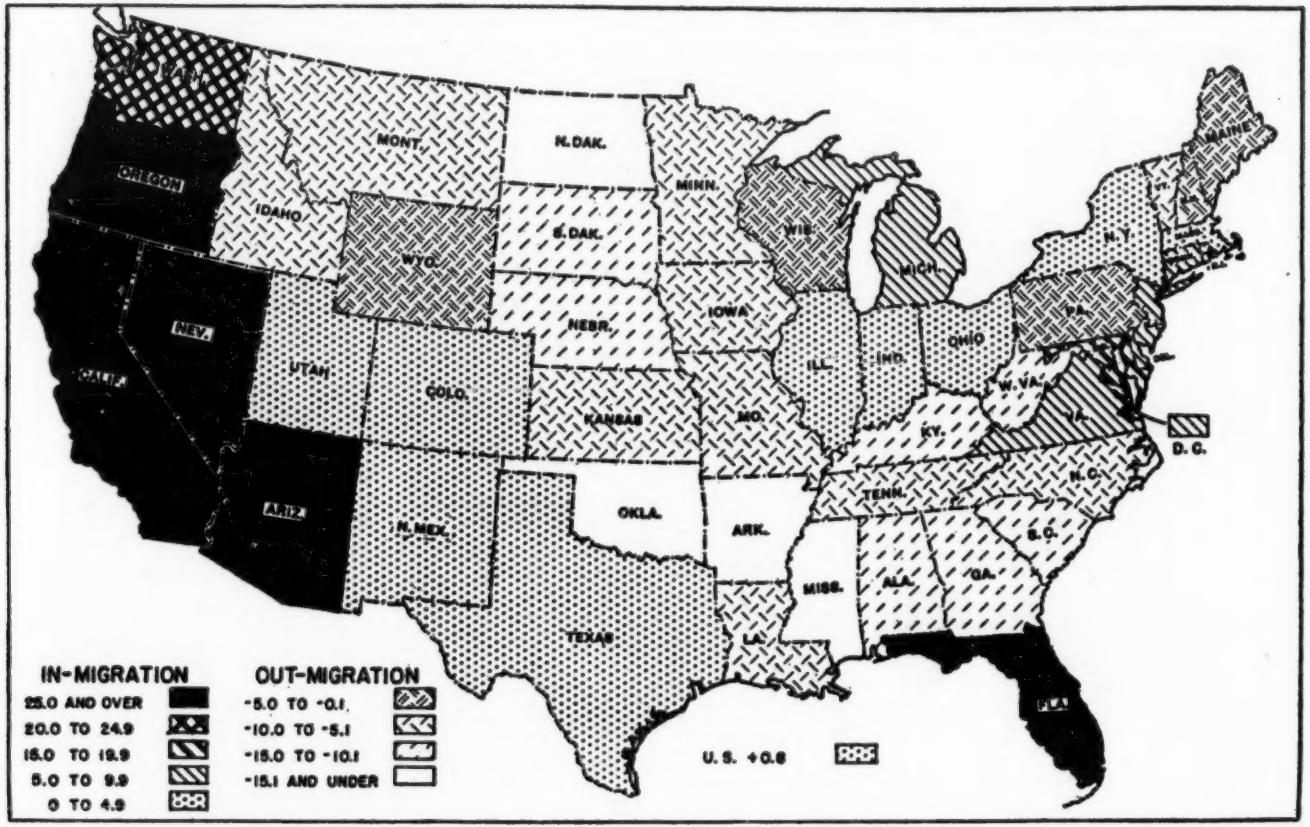


FIGURE 3. Net migration, 1940 to 1950, as per cent of 1940 population, by States

by states: (1) the excess of births over deaths[3] and (2) net migration, both as a percentage of the 1940 population [8]. Net migration is an estimate derived by subtraction and includes not only interstate migration but also international migration. (The latter amounted to about a million persons net for continental United States.)

All the States had a natural increase. The United States average was 13.7 per cent of the 1940 population. The percentage increase from this cause ranged from eight per cent for New York to 26 per cent for New Mexico. Twenty-six of the States had a net out-migration; but, in all but four of these, the natural increase was large enough to overcome this deficit. Areas of out-migration included: (1) Northern New England; (2) Pennsylvania; (3) the West North Central States plus Wisconsin; (4) The South except for Delaware, Maryland, the District of Columbia, Virginia, Florida, and Texas; (5) Montana, Idaho, and Wyoming. As a percentage of the 1940 State population, Florida, Arizona, Nevada, and the Pacific States had a net migratory gain of at least 20 per cent. California's net migration was more than one-third of its population at the beginning of the decade. Maryland's migratory gain was about 15 per cent; and 6 other States, all east of the Mississippi, gained by 5 to 10 per cent. Fifteen States had a net migration of less than 5 per cent, plus or minus. Ten States had migratory losses of 5 to 10 per cent, and seven had losses of 10 to 15 per cent. The latter were either in the South or in the Great Plains. Finally, Arkansas, Mississippi, North Dakota, and Oklahoma had a net migratory loss of more than 15 per cent.

Net civilian migration by States from April 1, 1940 to November 1, 1943, was estimated by Eldridge on the basis of war ration book registrations [9]. The amount of net migration for this period is compared with the civilian net migration for the whole decade in Figure 4. The differences should represent largely civilian net migration between 1943 and 1950 plus that portion of the civilian changes of State of residence between 1940 and 1950 that was made by persons who had been in the armed forces in 1943. The larger differences should be fairly reliable indicators of what happened after 1943.

Net civilian immigration from abroad was only 305,000 up to the middle of the war but 1,600,000 for the whole decade. A large part of all immigration has gone to New York. The comparative figures suggest

FIGURE 3. Net migration, 1940 to 1950, as per cent of 1940 population, by States

[3] These are births and deaths occurring to residents of the area regardless of migration status. Thus this figure does not give the change that would have occurred during the decade had the population of the state remained "closed."

that the following States had net gains from civilian migration in both parts of the decade: Massachusetts, New Jersey, Ohio, Indiana, Illinois, Michigan, Delaware, Florida, Arizona, Nevada, Washington, Oregon, and California. Relatively heavy gains in the 1943–1950 period seem to have occurred in Florida and Oregon. Connecticut and Maryland, which gained through migration up to the middle of the war, had little if any net gain thereafter. Important war plants in these States reduced operations after 1945, creating a considerable reduction in employment that was not fully offset by peacetime production. States where this situation had an even heavier impact were Rhode Island, the District of Columbia, Virginia, and Utah, all of which had smaller civilian migratory gains for the whole decade than for the first half. The large amount of new housing constructed in the Maryland and Virginia suburbs of Washington was an important factor in the postwar exodus from the capital.

Of the states that had had a net loss from civilian migration up to 1943, four had a net gain for the whole decade. These were New Hampshire, New York, Colorado, and New Mexico. About 10 states had a smaller net migratory loss for the period 1940 to 1950 than for the period 1940 to 1943, also suggesting a net gain from migration between 1943 and 1950. These were Maine, Vermont, Wisconsin, Minnesota, Iowa, South Dakota, Texas, Montana, Idaho, and Wyoming. Finally, 16 states evidently had net out-migration in both parts of the decade: Pennsylvania, Missouri, North Dakota, Nebraska, Kansas, West Virginia, North Carolina, South Carolina, Georgia, Kentucky, Tennessee, Alabama, Mississippi, Arkansas, Louisiana, and Oklahoma. Out-migration after the middle of the war seems to have been relatively heavy as compared with the earlier out-migration in the bloc of states comprising Georgia, Alabama, Mississippi, Louisiana, and Tennessee.

It should be emphasized that the statements concerning the period 1943 to 1950 are based on differences of differences and are thus subject to considerable error. The characterizations of Texas and Wyoming, for instance, may well be quite erroneous.

It has been pointed out by Shryock and Eldridge that the number of net migrants by states in a given period has had a high positive correlation with the number in another period, at least up to twenty years earlier or later [10]. The correlation coefficient of the present estimates for 1940 to 1950 with those for 1930 to 1940 is $+0.91$, despite the fact that economic conditions during the two decades were in marked contrast.

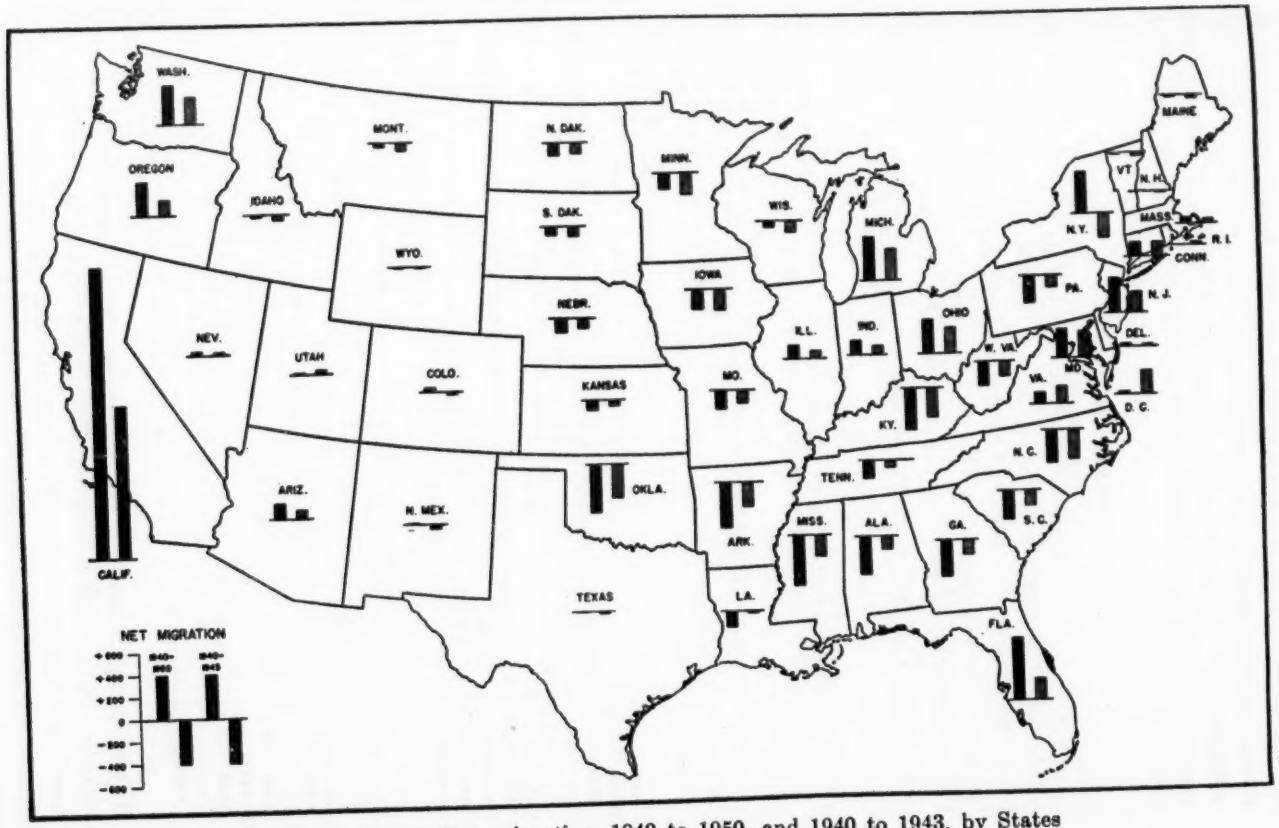


FIGURE 4. Net civilian migration, 1940 to 1950, and 1940 to 1943, by States

It will be possible to estimate net migration from 1940 to 1950 for cities, counties, standard economic areas, and other areas for which vital statistics can be compiled. When the final 1950 population figures become available, the Census Bureau intends to prepare estimates of this sort.

GROSS MIGRATION

The data of the foregoing section do not tell us anything about the total amount of migration, in-migration to a given area, out-migration from a given area, or streams of migration from one area to another. So far, very little information of this sort is available from the new census. For an indication of what happened during the 'forties, we must turn to the Current Population Survey, a national sample survey.

TABLE 3

MIGRATION STATUS AND TYPE OF MIGRATION FOR THE CIVILIAN POPULATION OF THE UNITED STATES: SELECTED PERIODS, 1940 TO 1950

Number

Period	Population 1 year old and over	Non-migrants	Migrants			Persons abroad at beginning of period
			Total	Intrastate	Interstate	
March 1949 to March 1950	146,864,000	138,124,000	8,250,000	4,360,000	3,889,000	491,000
April 1948 to April 1949	144,101,000	135,289,000	8,335,000	3,992,000	4,344,000	476,000
April 1947 to April 1948	141,698,000	132,228,000	9,008,000	4,638,000	4,370,000	462,000
April 1940 to April 1947	122,633,000*	96,565,000	25,469,000	13,081,000	12,388,000	599,000
			Per cent			
March 1949 to March 1950	100.0	94.0	5.6	3.0	2.6	0.3
April 1948 to April 1949	100.0	93.9	5.8	2.8	3.0	0.3
April 1947 to April 1948	100.0	93.3	6.4	3.3	3.1	0.3
April 1940 to April 1947	100.0	78.7	20.8	10.7	10.1	0.5

* Seven years old and over.

Table 3 compares some of the periods for which migration data have been collected through the current Population Survey. As has been explained elsewhere, it is not valid to compute the average annual migration by dividing the total reported for the period 1940 to 1947 by seven, Nor do we obtain any meaningful figures for the decade by adding the

absolute figures for the component periods [11]. It can be safely inferred from the available data, however, that internal migration took place at a higher annual rate during the war and reconversion periods than during the prewar period. There was also some falling off of civilian migration in the last two years of the decade.

Not very much is known as yet about the migration to and from specific parts of the United States. Some information for selected cities and counties is furnished by the Census of Congested Production Areas taken in 1944 and by pretests of the 1950 Census taken between 1947 and 1949. For 1941 to 1945 and 1940 to 1947, the movement to and from each of the four census regions was obtained in the Current Population Survey.

Data on movements to and from farms are available from both the Current Population Survey of the Bureau of the Census and from a national sample survey conducted annually by the Bureau of Agricultural Economics. The latter is chosen for discussion here because it provides an annual series back to 1940. Table 4 is condensed from the latest report. It may be noted that in addition to the well-known movement from farms, hundreds of thousands of persons move *to* farms every year.

TABLE 4

ANNUAL ESTIMATES OF THE FARM POPULATION, NATURAL INCREASE OCCURRING IN THE FARM POPULATION, AND NUMBER OF PERSONS MOVING TO AND FROM FARMS, FOR THE UNITED STATES: 1940 TO 1949

(Population in 000's)

Year	Farm population on January 1	Natural increase	Change through migration						
			Total	Civilian migration			Military migration		
				Total	To farms	From farms	Total	To farms	From farms
1949	27,776								
1948	27,440	479	−143	−123	1,178	−1,301	−20	−79	59
1947	27,550	483	−593	−640	914	−1,554	47	−69	116
1946	26,850	419	281	−266	1,077	−1,343	547	−159	706
1945	25,190	354	1,306	603	1,684	−1,081	703	−191	894
1944	25,521	345	−676	−476	817	−1,293	−200	−300	100
1943	26,659	355	−1,493	−988	994	−1,982	−505	−605	100
1942	29,048	390	−2,779	−1,920	819	−2,739	−859	−859	—
1941	29,988	417	−1,357	−1,146	814	−1,960	−211	−211	—
1940	30,269	400	−681	−606	690	−1,296	−75	−75	—

Source: U. S. Bureau of Agricultural Economics. *Farm population estimates, January, 1949.* June 1949.

It is usually the case that gross migration for an area is much larger than net migration and that short-distance mobility is much larger in

volume than long-distance migration. These generalizations fit what we know about the 'forties rather well. The most important exceptions are on the Pacific Coast. The amount of net in-migration for many areas there was not very much smaller than the amount of gross in-migration. Furthermore, many migrants to the West Coast came from hundreds of miles away.

DIFFERENTIALS IN POPULATION REDISTRIBUTION

It is well known that migration is selective of particular population groups. The Current Population Survey of April 1947, for instance indicated a disproportionately heavy movement of Negroes from southern farms since 1940. In addition to color, these surveys have related-migration status to sex, age, marital status, duration of marital status, family status, educational attainment, employment status, major occupation group, and income. These data are presented in various publications of Census Bureau Series P-20, P-50, and P-60.

From a preliminary sample of the 1950 Census, a few data on migration status are being published for the United States, the 10 largest states, and the 55 standard metropolitan areas and 36 cities of 250,000 or more in 1940 [12]. Distributions for the United States from this source may be compared with those from the Current Population Survey: The differences are too large to be explained by the small variation in dates or by sampling error. There is by now considerable evidence

Residence in 1949	April 1949 to April 1950, 1950 Census	March 1949 to March 1950, C.P.S.
Same house	82.4	80.9
Different house, same county	11.3	13.1
Different county or abroad	6.3	6.0

that the better trained, experienced enumerators of the Current Population Survey obtain a more nearly complete identification of the unemployed and the marginal members of the labor force than do the decennial census enumerators. It seems likely that they would also get a more nearly accurate classification by other items defined by answers to a series of questions, such as migration status, but we have no direct evidence on this point. Further information on accuracy should be obtained later from the Postenumerative Survey of the 1950 Census and from other studies.

Among the 46 Standard Metropolitan Areas for which these figures have been published to date, the percentage of those reporting their 1949 residence who were migrants (or immigrants from abroad) ranged from 1 per cent to 16 per cent. This percentage varied roughly with the metropolitan area's rate of population increase during the whole dec-

TABLE 5

PER CENT DISTRIBUTION BY RESIDENCE IN 1949 OF THE POPULATION 1 TO 20 YEARS OLD, BY AGE, FOR THE UNITED STATES: 1950

Residence in 1949	Total 1 to 20 years old	1 to 4 years	5 to 9 years	10 to 13 years	14 to 17 years	18 to 20 years
Total	100.0	100.0	100.0	100.0	100.0	100.0
Same house as in 1950	80.8	77.2	82.5	85.8	86.1	70.5
Different house, same county	11.4	13.8	10.8	9.2	8.4	15.2
Different county, same state	3.3	3.9	2.9	2.2	2.4	5.5
Different state or abroad	3.2	4.0	2.6	1.8	2.1	6.6
Residence not reported	1.3	1.2	1.2	1.0	1.1	2.3

ade. At the extremes, it was below 3 per cent in Johnstown, Scranton, Wilkes-Barre-Hazleton, and Worcester and above 10 per cent in Denver, Miami, and Washington. For metropolitan areas containing two or more counties, this movement includes migration within the metropolitan area.

For use at the Midcentury White House Conference on Children and Youth, some preliminary sample data on migration from the 1950 Census have been published for the population under 21 years old [13]. Table 5 also is derived from this source. Of the 50,800,000 children and youths 1 to 20 years old, 9,100,000 had lived in a different house year earlier. Of these, 3,300,000 had lived in a different county and are thus classified as migrants. Those coming from a different state or from abroad numbered 1,600,000. The age groups of Table 5 are the most detailed for which census migration data have been presented. They reflect some interesting aspects of the family cycle. Note the decline in mobility from "1 to 4" years to "5 to 9" years and the further decline to "10 to 13." There is little change at "14 to 17" and a sharp rise at "18 to 20." Many families move to the suburbs when the first child is old enough to play outdoors. When children are in high school, families are relatively settled and unlikely to move. Up to about age 18, most children move with their parents. Thereafter they often move on their own to attend college, serve in the armed forces, take jobs, or get

married. Presumably the age differentials examined would be even sharper if the data were restricted to first children. Similar variations of migration status with age were observed for the urban, rural-nonfarm, and rural-farm populations of the United States and for each region.

THE 1950 PROGRAM OF MIGRATION STATISTICS

Considerably more light should be shed on the ways in which population redistribution has been affected if the Census Bureau's contemplated program in this field can be carried out. The estimates of net migration for the period 1940 to 1950 have already been mentioned. Revised intercensal population estimates as of each year in that decade are being prepared for states and the larger cities and standard metropolitan areas. These should indicate the fluctuations of growth in various parts of the country as they were affected by extraordinary military and economic events.

From the migration questions on the 1950 Population schedule, many facts are potentially available about mobility in the year preceding the enumeration. Migration status will be published for counties, cities of 10,000 or more, and census tracts. The tables showing migration from each state to every other state that were published in 1940 will again be prepared. The new state economic areas will also figure importantly in the reports on internal migration. Other new emphases will be on the short-distance movements from one house to another within the same county and, for the larger streams of migration, on social and economic characteristics cross-classified by age. Outlines of the data to be tabulated may be obtained by writing to the Population and Housing Division, U. S. Bureau of the Census, Washington 25, D. C.

SUMMARY

The record-breaking amount of population growth between 1940 and 1950 was very unevenly distributed over the Nation. World War II and its prosperous aftermath accelerated the concentration of people in metropolitan areas and around our sea coasts, as almost half of the counties lost population. At the same time, the decentralization within metropolitan areas also continued. Differentials in the excess of births over deaths have contributed to population redistribution; but migration, and notably internal migration, has been more important in this respect.

Some of the survey data concerning internal migration that have been collected during the 'forties are summarized in the accompanying tables and maps, along with the first migration statistics from the 1950 Census. Within the next few years, it is hoped that this source will give us a much fuller knowledge of the dynamics of population redistribution.

REFERENCES

[1] U. S. Bureau of the Census. *Preliminary counts.* Series PC-3, No. 4. "Population of Counties: April 1, 1950." November 5, 1950.

[2] U. S. Bureau of the Census. *Preliminary counts.* Series PC-3. No. 3. "Population of Standard Metropolitan Areas: April 1, 1950." November 5, 1950.

[3] Bogue, Donald J. "Changes in Population Distribution since 1940," *American Journal of Sociology,* 56(1): 43–57. July 1950.

[4] U. S. Bureau of the Census. *Preliminary counts.* Series PC-3, No. 7. "Population of State Economic Areas: April 1, 1950." December 27, 1950.

[5] U. S. Bureau of the Census. *State Economic Areas,* by Donald J. Bogue. Government Printing Office, Washington, 1951.

[6] Beck, P. G. and Forster, M. C. *Six rural problem areas.* Federal Emergency Relief Administration, Research Monograph I, Washington, 1935: 137.

[7] See, for example: Thompson, Warren S., and Bogue, Donald J. "Subregional Migration as an Area of Research." *Social Forces,* 27 (4): 395. May 1949.

[8] U. S. Bureau of the Census. *Current population reports. Population estimates.* Series P-25, No. 47. "Provisional Estimates of the Population of Regions, Divisions, and States: July 1, 1940 to 1949." March 9, 1951.

[9] U. S. Bureau of the Census. *Population—Special Reports.* Series P-44, No. 17. "Interstate Migration and Other Population Changes: 1940 to 1943," by Hope Tisdale Eldridge, August 28, 1944.

[10] Shryock, Henry S., Jr., and Eldridge, Hope Tisdale. "Internal Migration in Peace and War." *American Sociological Review,* 12(1): 27–39. February 1947.

[11] Shryock, Henry S., Jr. "Wartime Shifts of the Civilian Population." *Milbank Memorial Fund Quarterly,* 25(3): 270. July 1947.

[12] U. S. Bureau of the Census. *Preliminary counts.* Series PC-7, No. 1 "General Characteristics of the Population of the United States: April 1, 1950." February 25, 1951; . . . Series PC-6, Nos. 1–10. "Characteristics of the Population of [State]." March 20–April 10, 1951; Series PC-5, Nos. 1-55. "Characteristics of the Population of the . . . Standard Metropolitan Area: April 1, 1950." March 28, 1951 to date.

[13] U. S. Bureau of the Census. *Current Population reports. Population characteristics.* Series P-20, No. 32 "Children and Youth: 1950." December 4, 1950.

CHANGES IN INCOME DISTRIBUTION IN THE UNITED STATES

HERMAN P. MILLER
Bureau of the Census

BETWEEN 1936 and 1948 aggregate consumer income in the United States tripled. Available evidence indicates that this increase was accompanied by a marked change in the distribution of income among American families. In 1936 about half of the families and individuals[1] received incomes of less than $1000 and only 3 per cent received incomes of $5000 or more. In 1948 only about one-tenth of the families and individuals had incomes under $1000 while the proportion with incomes of $5000 or more increased eightfold to about 25 per cent.

TABLE 1

CUMULATIVE PER CENT DISTRIBUTION OF FAMILIES AND INDIVIDUALS AND OF INCOME RECEIVED BY FAMILIES AND INDIVIDUALS, BY TOTAL MONEY INCOME LEVEL, FOR THE UNITED STATES: 1935–36 AND 1948

Total money income level	Families and individuals		Income received by families and individuals	
	1948*	1935–36†	1948†	1935–36†
Under $ 1,000	9.8	48.9	1.1	18.1
Under $ 2,000	26.5	81.0	7.0	48.2
Under $ 3,000	42.8	92.3	16.7	66.0
Under $ 4,000	60.5	95.6	31.4	73.6
Under $ 5,000	74.1	97.0	45.7	77.8
Under $10,000	95.3	99.1	79.6	87.1
Total	100.0	100.0	100.0	100.0

* *The Economic Report of the President*, January 1950, p. 140.
† Unpublished data underlying the distributions published by the Council of Economic Advisers.

In terms of 1948 dollars the average income of the lowest fifth of the families and individuals increased by 67 per cent between 1936 and 1948 and that of the next lowest fifth increased by 93 per cent. At the opposite end of the income scale, the income of the wealthiest fifth increased by only 40 per cent.[2] The greater relative gains of the lower income groups suggest that income inequality decreased during this period.

The changes in the share of the income received by each quintile are

[1] The term "family" refers to a group of two or more persons related by blood, marriage or adoption and residing together. The term "individual" (or "unrelated individual" in Census terminology) refers to persons other than inmates of institutions, who are not living with any relatives.
[2] *The Economic Report of the President*, January 1950, p. 97.

even more indicative of a diminution of inequality in the distribution of income. Between 1936 and 1948 the share of the income received by the lowest 40 per cent increased from about 13 per cent to 15 per cent and the share received by the middle 40 per cent increased from 34 per cent to 38 per cent. In contrast, the share of the income received by the highest fifth decreased from 53 per cent to 47 per cent.[3]

Although these data suggest a diminution of inequality, they are subject to many qualifications. The data were obtained from two different field surveys of income which differed in such important respects as sample design, survey technique, and the training of the interviewers. This affects their comparability. In addition, there are differences in the adjustments which have been made to the raw survey data which may well be the most important source of incomparability in the data. These differences are not minor statistical discrepancies. The adjustments made to the data for 1948 involved about one-fifth of the aggregate income for that year. The adjustments which have been made in the interest of comparability cannot be guaranteed to achieve their purpose fully.

For this reason, it may be of some interest to examine income statistics covering a prewar year (1939) and a postwar year (1949) which have been collected by the Bureau of the Census, as a check on the conclusions. Although the Census data also contain elements of incomparability, they are alike in most basic respects and can more readily be compared than the data shown above. The most serious shortcoming of the comparison made here is that in the absence of information on total income for 1939, the data had to be restricted to wage-earner families and individuals (i.e., those having wage or salary income and less than $50 of nonwage income). Since nearly one-half of all families and individuals in the United States in 1939 and 1949 were in this category, however, the data reflect income changes for a large segment of the population.

A per cent distribution of wage-earner families and individuals by wage or salary income level in 1939 and 1949 is shown in the table below. This table shows clearly the marked change in the pattern of income distribution since 1939. In 1939 the average (median) wage-earner family or individual had an income of only $1300. By 1949 this figure had more than doubled. In 1939 less than 2 per cent of the wage-earner families and individuals had incomes of $5000 or more and over one-third of them were concentrated at the low end of the income scale with incomes of less than $1000. By 1949 the proportion having incomes

[3] *Ibid.*, p. 97.

of $5000 or more had increased tenfold to 19 per cent, whereas the proportion with incomes under $1000 was reduced to 6 per cent.

TABLE 2

CUMULATIVE PER CENT DISTRIBUTION OF WAGE-EARNER FAMILIES AND INDIVIDUALS WITHOUT NONWAGE INCOME, AND OF WAGE OR SALARY INCOME RECEIVED BY THESE FAMILIES AND INDIVIDUALS, BY WAGE OR SALARY INCOME LEVEL, FOR THE UNITED STATES 1939 AND 1949

Wage or salary income level	Families and individuals*		Income received by families and individuals†	
	1949	1939	1949	1939
Under $ 500	2.4	16.4	0.2	2.6
Under $1,000	6.4	37.2	1.0	12.4
Under $1,500	12.4	57.7	2.9	28.5
Under $2,000	20.0	74.7	6.3	47.1
Under $2,500	31.7	85.5	13.2	62.4
Under $3,000	44.1	91.1	22.1	72.0
Under $5,000	81.2	98.4	60.7	90.3
Total	100.0	100.0	100.0	100.0
Median income	$3,323	$1,312	—	—

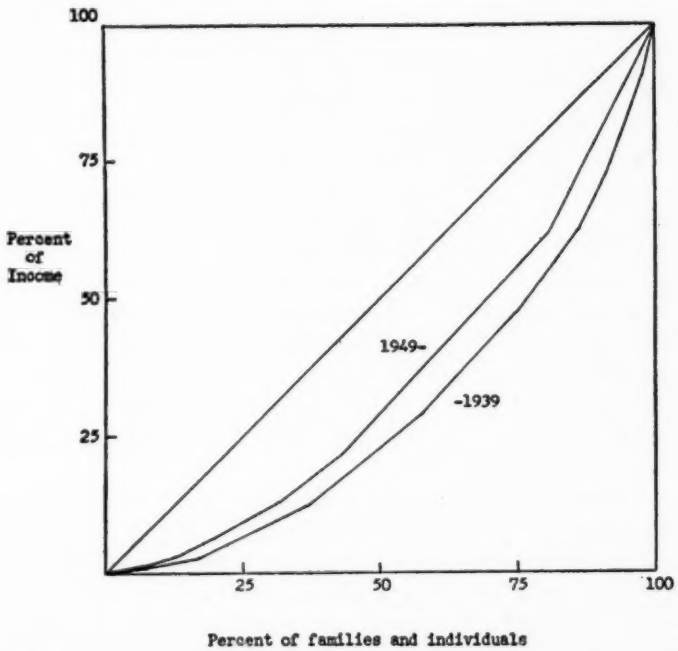
* Department of Commerce, Bureau of the Census, *Income of Families and Persons in the United States: 1949*, Series P-60, No. 7, p. 27.

† Estimates derived from distributions of families and individuals by wage or salary income levels (see text).

The measurement of the change in income concentration requires a knowledge of the distribution of aggregate income. Unfortunately, due to the absence of tabulations of aggregates for either year, or of a distribution of the 1939 data by more detailed income levels, only a rough approximation of the distribution of aggregates for these years can be made. In order to maintain comparability, the same class intervals were used to compute the aggregates for 1939 and 1949 despite the fact that more detailed class intervals were available for 1949. The aggregates for each year were estimated in the following manner. For each income level below $5000, the number of families and individuals was multiplied by the midpoint of the class interval. The averages used for the $5000 and over interval were $9000 for 1939 and $8000 for 1949. These averages were obtained from an analysis of income tax data. A check on the distribution of aggregate income for 1949 was made by recomputing it from a more detailed distribution divided into 21 class intervals with an estimate for the open-end interval ($10,000 and over) based on a Pareto Curve. There was no significant difference between the two distributions.

The change in the distribution of families and individuals by income levels was accompanied by a marked change in the distribution of ag-

gregate income. The accompanying chart shows that there was an appreciable change in the distribution of income among wage-earner families and individuals in the direction of greater equality between 1939 and 1949. In 1939 approximately 11 per cent of all the income received by wage-earner families and individuals went to those who



Lorenz Curves Showing the Distribution of Wage or Salary Income Among Wage-Earner Families and Individuals, for the United States: 1939 and 1949

were in the lowest third of the income distribution; in 1949 the lowest third received about 15 per cent of the aggregate income. At the opposite end of the income scale, 62 per cent of the income went to the wealthiest third of the families and individuals in 1939, whereas only 54 per cent was received by this group in 1949. Although the changes indicated by the Lorenz curves appear small, they signify great relative changes in the distribution of income.[4] For example, the increase of 4 percentage points (from 11 to 15 per cent) in the share of the income received by the lowest third represents an increase of nearly 40 per cent in the relative share ot the income received by this group.

[4] A discussion of factors related to these changes in income distribution appears in an article by the author in *The Review of Economics and Statistics*, August 1951.

RESULTS OF ALTERNATIVE STATISTICAL TREATMENTS OF SAMPLE SURVEY DATA*

L. R. KLEIN† and J. N. MORGAN‡

1. INTRODUCTION

A CAREFUL formulation of the probability and economic model which underlies observations is indispensable for the choice of correct statistical procedures to be followed in drawing econometric inferences from samples. The sample design, the methods of collecting data, and underlying economic behavior will all contribute to the formulation of the model. The study of data collected in consumer surveys[1] has convinced us that one cannot proceed simply by the application of conventional statistical methods in the estimation of economic relationships because of the existence of some basic difficulties which we classify as follows: (1) weighting of observations, (2) heteroscedasticity, (3) nonlinearities, (4) the choice of alternative economic concepts, (5) errors of observation. In this paper, we shall describe some empirical calculations that deal with the first four difficulties. The fifth is reserved for future study.

Observations may be weighted because of the use of differential sampling rates and because of differential response rates to survey questions in the different strata of the population. High rent (a correlate of high income) areas are deliberately over-sampled because they show more variability in respect to certain economic variables than other strata, and conversely low rent areas are under-sampled. The strata also have different response rates. The weights finally used are combinations of sampling and response rates which attempt to give each stratum the same relative importance in the sample that it has in the population; i.e., they are essentially inversely proportional to the net (effective) sampling rate.

The heteroscedasticity which leads to the adoption of differential sampling rates must also be taken into account in the estimation of parameters in equations expressing economic behavior. In these equations, the variance of the so-called disturbances is a fundamental parameter to be estimated, and this variance is not uniform over some

* We acknowledge helpful criticism from our colleagues P. S. Dwyer, R. Goodman and G. Katona. The research reported in this paper was carried out under a grant from the Carnegie Corporation.

† National Bureau of Economic Research and Survey Research Center, University of Michigan.

‡ Survey Research Center, University of Michigan.

[1] Specifically, we confine ourselves to *The Surveys of Consumer Finances* conducted by the Survey Research Center, University of Michigan, for the Board of Governors of the Federal Reserve System.

strategic classifications. For example, if we consider a relation between household savings (S) and household income (Y)

$$S = f(Y) + u, \tag{1.1}$$

we shall find that the variance of u grows with Y. This point has been often observed in survey data[2] and is, in fact, a reason for oversampling high-rent areas in the Surveys of Consumer Finances. Heteroscedasticity is a product of economic behavior, while oversampling and weighting of observations are statistical devices. The two types of considerations must be appropriately combined for the problem of the estimation of parameters of equation (1.1) from sample observations.

A simple procedure which is often followed is to make a linear correlation between mean savings and mean income over a number of income classes to estimate (1.1). Aside from the questions of weighting and heteroscedasticity, there is a question as to whether additional variables nonlinearly related would lead to superior estimates. Without any grouping of observations beyond the microeconomic spending unit, the household, we shall look for nonlinearities of two types in the savings equation. One type of nonlinearity is described by allowing the marginal propensity to save, $\delta S/\delta Y$, to vary with income. Our first approximation will be a simple parabola in Y. In addition to income we use beginning-of-year liquid asset holdings (L) as a variable influencing savings behavior. The second type of nonlinearity is described by permitting the marginal propensity to save to vary with L or the marginal asset effect to vary with Y. We arrive at this hypothetical formulation through considerations that place the strongest inverse relationship between liquid asset holdings and saving in the lowest income groups. The actual treatment in our study will be approximated by using L as a variable in an equation, linear in parameters,

$$S = \alpha_0 + \alpha_1 Y + \alpha_2 Y^2 + \alpha_3 L + \alpha_4 LY + u. \tag{1.2}$$

A simple model leading to (1.2) is composed of two separate equations.

$$S = \beta_0 + \beta_1 Y + \beta_2 Y^2 + \beta_3 L + v_1 \tag{1.3}$$

$$\beta_1 + 2\beta_2 Y = \gamma_0 + \gamma_1 L + v_2. \tag{1.4}$$

---

[2] See H. Mendershausen, "The Relationship between Incomes and Savings of American Metropolitan Families," *The American Economic Review*, Vol. 29, (1939). pp. 521–537, esp. p. 525.

Equation (1.3) is a savings equation nonlinear in income and dependent on liquid assets. Equation (1.4) makes the marginal propensity to save a linear function of liquid assets. This equation allows for the possibility that, in a cross section sample, high asset holders are those who have saved much in the past and will continue to do so, at the margin, in the future. The two equations together yield, when we substitute into (1.3) the value of β_1 found from (1.4),

$$(1.5) \qquad S = \beta_0 + \gamma_0 Y - \beta_2 Y^2 + \beta_3 L + \gamma_1 LY + v_1 + v_2 Y.$$

This derived equation has the same general form as (1.2). There is no inconsistency in the disturbances of the two equations since v_1 is actually of the form vY—the problem of heteroscedasticity already mentioned. An interesting property of (1.5) is that the coefficient of Y^2 is opposite in sign to the coefficient of Y^2 in (1.3).

The variables in equation (1.2) have not yet been precisely defined. There are a variety of different concepts that can be used to describe savings, income, or liquid assets depending on the treatment accorded such things as purchase of durable consumer goods, depreciation of consumer assets, nonmoney incomes, capital gains, other income transfers, securities other than U. S. government bonds, etc. As far as income and liquid assets are concerned, we have used the definitions adopted in the Surveys of Consumer Finances;[3] thus, income does not include nonmoney income, capital gains, or many types of transfers. Our variable Y is disposable income; i.e. federal personal income taxes have been estimated for each sampled spending unit by the Survey Research Center and subtracted from total money income to get disposable income. Liquid assets include only U. S. government bonds, bank deposits, postal savings accounts, shares in savings and loan associations, and shares in credit unions. For savings, we have made parallel calculations for two different concepts, one concept being that used in the Surveys of Consumer Finances[4] and the other being the same variable plus expenditures on consumers' durable goods, written as S'. The variable S' accords similar treatment to both consumer durable and residential housing outlays, whereas S treats housing but not purchases of other durables as savings.

The use of weights because of differential sampling and response

---

[3] See, "1950 Survey of Consumer Finances", Parts III and V, *Federal Reserve Bulletin*, August and December, 1950.

[4] See, "1950 Survey of Consumer Finances", Part IV, *Federal Reserve Bulletin* November, 1950.

rates by strata is applicable to the spending unit as a whole, i.e. every variable from a given spending unit is given the same weight. Some variables are determined with high precision from a given spending unit, while others are determined with much lower precision. Age of head, family size, and home ownership are quite precise; while savings, income, assets, and debts are subject to more sizeable errors of observation. Even among the latter variables, however, there are varying degrees of accuracy. Income is probably more accurately observed than savings. Memory error is present in getting beginning-of-year liquid assets[5] and may make for less precision in the observations for L than for S and Y.[6] We have used weights that are uniform for all variables pertaining to a given spending unit and have not gone into the problem of attaching different weights to different variables for the same spending unit.

2. COMMENTS ON METHOD AND SAMPLE

The variables described in the preceding section, S, S', Y and L are not the only factors that will appear in a definitive equation of savings behavior, but extensive studies of other variables have shown that the major effects can be accounted for by a nonlinear relation of the form (1.2). Since the empirical work in this paper involves estimates of several different equations from sample observations of individual households, we have tried to keep the work at manageable levels by including only the most important variables. We realize that the equation can be improved by the use of variables that we have studied in other connections, namely, income-change, income-expectations, age of head of spending unit, home ownership, city size, and spending-unit size. For the most part these other variables do not affect the coefficients of Y, Y^2, L, or LY and essentially contribute to refinement rather than basic characterization of the equations. To a large extent these additional variables, except income-change, affect the level of the savings function without greatly altering our estimates of $\delta S/\delta Y$ or $\delta S/\delta L$. Further refinements might consist of classifying housing expenditures as consumer outlay rather than saving or drawing

[5] Estimates of savings equations have been made from a special reinterview sample of urban spending units in which beginning-of-year liquid assets were taken from the previous year's survey. Results similar to those of the current paper were found, although the marginal asset effects appear to be significantly larger from the reinterview data.

[6] One of the components of S is $L_1 - L$ (L_1 = end-of-year liquid assets), but other components are probably less seriously affected by memory error. Furthermore, in some cases asset change is reliably estimated, while the levels are not.

a distinction between contractual (insurance premiums, mortgage payment, other debt repayment) and other types of saving.[7]

The mathematical form of the savings function used in this paper may seem restrictive, but we find that it gives a reasonably good approximation to the observed data. We have experimented extensively with other forms and do not find significant differences that would affect the conclusions we eventually draw from the results of this paper. Another equation considered elsewhere involves the use of a logarithmic transformation of income in order to approximate, with fewer parameters, the essential nonlinearity of the savings-income relation. We have also studied relations between variables within asset or income groupings. The results of the other studies are here used principally to check our findings from the particular type of equation we have assumed for this paper.

Our sample consists of 2281 spending units drawn from the 1949 Survey of Consumer Finances. The 2281 observations exclude 1229 spending units in the 1949 Survey. We have excluded this large number of cases for reasons dependent on the objective of our present analysis, and for reasons associated with different research purposes. We feel that the sample of 2281 is both adequate and convenient for our present investigations. We also have an idea of the nature of the main bias introduced by the selection of our sample. The bias undoubtedly works in the direction of decreasing the estimate of the variance of the disturbances associated with our savings equations.

The two major exclusions from our sample consist of 430 farmers and 383 business owners; thus, we are admittedly trying to infer savings behavior in the nonfarm, nonbusiness sector of the household population. Accounting problems are exceedingly complex for farmers and businessmen; therefore, we feel that it is wiser to study their behavior separately. Moreover, these complexities leave us in great doubt as to the appropriate concepts of savings, income, and liquid asset holdings for these spending units acting purely as consumers, if such action is conceivable. The prevalence of nonmoney income and inventory fluctuations among farmers also presents a serious problem. Finally, separate investigations of savings-income regressions for farm-

[7] Calculations using these numerous additional variables are reported in the following places: L. R. Klein, "Assets, Debts, and Economic Behavior" to be published in a forthcoming volume of *Studies in Income and Wealth*, National Bureau of Economic Research; L. R. Klein, "Sample Surveys of Households: A New Tool in Econometrics," paper presented at the December, 1950, meetings of the Econometric Society in Chicago, Illinois; J. N. Morgan, "Consumer Savings," paper presented at the September, 1950, meetings of the Econometric Society in Cambridge, Massachusetts, abstract in *Econometrica*, Vol. 19, 1951, pp. 63–64.

ers and businessmen indicate that their savings behavior differs significantly from that of the rest of the population.

Another large group of omissions consists of 139 spending units whose saving was not completely ascertained in the 1949 Survey. The Survey Research Center has assigned a value to each missing component of savings for these spending units by methods resembling the Yates missing plot technique, but we prefer to use only definitely reported observations which amounts to assigning to each of the excluded cases the mean values of the sample without including them in the number of degrees of freedom. In other studies we have found that estimated equations using assigned values hardly differ from those omitting these observations. In addition we have, for the same reasons, omitted 50 observations for which saving was ascertained but liquid assets were not, 37 observations for which durable goods expenditures were not ascertained, 14 observations for which disposable income was not ascertained, and 6 observations for which the fact of business ownership was not ascertained.

The following cases are omitted for reasons extraneous to this particular study:

87 for which income change was not ascertained;
17 for which contractual savings either were contractual in form but not in substance or were not ascertained;
7 for which residual expenditures (income less savings less durable goods expenditures) were less than $200;
2 for which home ownership was not ascertained;
55 of doubtful reliability or involving unusual circumstances such as extremely large medical expenses or other emergency outlays;
2 of very high disposable income, outlying observations. The incomes were $36,920; and $72,813.

Other studies involving a larger number of variables are being made with this sample, hence, these extraneous reasons for excluding 170 observations. The numbers given for each type of excluded case depend on the order in which we used variables for picking out interviews to be omitted, since some spending units are excluded for more than one reason.

An indication of the nature of our sample is given by comparing the (relative) frequency distributions of savings, disposable income, and liquid assets for excluded nonfarm, nonbusiness observations with the corresponding distributions for the observations actually used in our calculations. The *unweighted* distributions are given in the following table:

PERCENTAGE DISTRIBUTIONS OF SAVINGS, DISPOSABLE INCOME AND LIQUID ASSETS
NONFARM, NONBUSINESS; 1948

(units rounded to nearest dollar)

Savings			Disposable Income			Liquid Assets		
	Included	Excluded		Included	Excluded		Included	Excluded
−$1000 to −500	11%	26%	$ 0 to 999	10%	14%	$ 0	26%	27%
− 499 to −100	13	8	1000 to 1999	20	17	1 to 199	14	16
− 99 to 1	7	3	2000 to 2999	25	23	200 to 499	14	10
0	6	6	3000 to 3999	23	18	500 to 999	12	8
1 to 99	16	16	4000 to 4999	13	9	1000 to 1999	13	10
100 to 199	9	8	5000 to 7499	7	10	2000 to 2999	6	6
200 to 499	17	12	7500 to 9999	1	5	3000 to 4999	7	8
500 to 999	11	8	10,000 to ∞	1	4	5000 to 9999	5	8
1000 to 1999	7	8	Negative	—	0	10,000 to 24,999	3	6
2000 to ∞	3	5				25,000 to ∞	0	1
	100%	100%		100%	100%		100%	100%

Source: 1949 Survey of Consumer Finances.

The distributions of excluded cases by each of the three variables agrees well with the distributions of included cases except for the fact that we have excluded a disproportionately large number of negative savers. Extreme negative savers are often extraordinary cases, and for many of them we have felt that the survey data available were inadequate to account for the exceptional circumstances. Our feeling is that our results are applicable to the nonfarm nonbusiness population of the United States, and the fact that the estimated equations are not much affected by the use of assigned values for missing variables reinforces our belief in the applicability of the findings to this population. We do suspect, however, that any bias which may have occurred as a result of the selection process works in the direction of decreasing the variability about our equations without appreciably affecting the coefficients. This is especially true because of the exclusion of a relatively large number of dissavers.

3. EMPIRICAL FINDINGS

The first model we shall examine is that of a savings function which is nonlinear in income and dependent also on liquid asset holdings.

$$S = \alpha_0 + \alpha_1 Y + \alpha_2 Y^2 + \alpha_3 L + u. \tag{3.1}$$

The importance of Y as a variable in the savings function and the

general magnitude of α_1 are widely known from other studies with different samples. Our first investigations are thus designed to throw some light on the statistical significance and size of α_2 and α_3. Policies affecting income redistribution and the money supply are evidently dependent on these two parameters if such policies are to realize their usual objectives. At the same time we are estimating (3.1), we want also to see whether there is any reason to prefer

$$S' = \alpha_0' + \alpha_1' Y + \alpha_2' Y^2 + \alpha_3' L + u' \tag{3.1'}$$

to (3.1) as a description of the decision process of households, where $S' = S$ plus durable goods expenditures.

In the probability model, we make the following assumptions:[8]

1. Y and L are predetermined variables.
2. $u = vY$ or $u' = v'Y$.
3. v (v') is normally distributed with zero mean and variance σ^2, $(\sigma')^2$.

The classification of Y and L as predetermined variables in cross-section samples of the type we have means that we regard the joint marginal distribution of these two variables to be independent of the parameters of (3.1) or (3.1'). Savings in any given year make a negligible contribution to the income stream derived from earning assets. For this reason we posit a one-way influence from income to savings and not vice versa. Most persons in the nonfarm nonbusiness sector of the population accept their income level of any single year as a given variable and adjust their expenditures and savings to it. Beginning-of-year liquid assets cannot be influenced directly by subsequent savings; hence the time lag is relevant in classifying L as predetermined. In equations (3.1) and (3.1') liquid assets are a predetermined variable affecting savings. Later we shall study a model in which liquid assets affect both the level of savings and the marginal propensity to save. Ignoring the problem of weighting and taking the above assumptions as valid, the correct procedure is to choose estimates of α_i and $\sigma(\alpha_i'$ and $\alpha')$ which minimize $\sum_1^N v^2$ ($\sum_1^N v'^2$), where N = sample size. An interesting point to be cleared up by our calculations is a determination of the empirical difference in the estimates obtained by minimizing $\sum u^2$ or $\sum (u')^2$ instead of $\sum v^2$ or $\sum (v')^2$.[9]

[8] Essentially this formulation defines new disturbances v and v' which have variances independent of Y. In a more refined model we would probably improve the formulation by allowing σ_u or $\sigma_{u'}$ to depend on both Y and L. The refinement is suggested by empirical study of the residuals from savings-income regressions.

[9] This formulation is equivalent to some versions of the problem of nonadditive disturbances. See L. Hurwicz, "Systems with Nonadditive Disturbances" and H. Rubin "Note on Random Coefficients," *Statistical Inference in Dynamic Economic Models*, ed. by T. C. Koopmans, Wiley, N. Y., 1950.

Equations (3.1) and (3.1′) can be rewritten as

$$(3.2) \quad S/Y = \alpha_0(1/Y) + \alpha_1 + \alpha_2 Y + \alpha_3(L/Y) + v,$$

$$(3.2') \quad S'/Y = \alpha_0'(1/Y) + \alpha_1' + \alpha_2' Y + \alpha_3'(L/Y) + v'.$$

The ordinary least-squares regressions of S/Y or S'/Y on $1/Y$, Y, and L/Y will provide estimates of the unknown parameters which are consistent with the assumptions underlying the model. On the other hand the regressions of S or S' on Y, Y^2, and L do not in general lead to identical parameter estimates nor are they consistent with the underlying assumptions.

In the David-Neyman statement of the Markoff Theorem on least squares the variances of the random variables are assumed to satisfy

$$\sigma_i^2 = \frac{\sigma^2}{P_i}$$

where the P_i are known positive constants.[10] The least-squares estimation equations for parameters in a linear relation use the P_i as weights. In our model the P_i are replaced by $1/Y^2$. We get the same estimation equations whether we form moments weighted by $1/Y^2$ or first divide all variables by Y and then form unweighted moments of the transformed variables, i.e., those in (3.2) and (3.2′).

Our calculations show for $N = 2281$

$$(3.1.e) \quad S = -234 + 0.12Y + 0.0000032Y^2 + 0.0057L + u,$$
$$\qquad\qquad (39)\ (0.016)\ (0.0000011)\ \ (0.0045)$$
$$R = 0.40 \qquad \text{est. } \sigma_u = \$756$$

$$(3.1'.e) \quad S' = -315 + 0.25Y + 0.00000076Y^2 + 0.010L + u',$$
$$\qquad\qquad (36)\ (0.014)\ (0.00000098)\ \ (0.0041)$$
$$R = 0.61 \qquad \text{est. } \sigma_{u'} = \$693$$

$$(3.2.e) \quad S/Y = -152(1/Y) + 0.13 + 0.0000054Y - 0.093(L/Y) + v,$$
$$\qquad\qquad (15)\ \ \ (0.017)\ (0.0000034)\ \ (0.0035)$$
$$R = 0.60 \qquad \text{est. } \sigma_v = 0.28$$

$$(3.2'.e) \quad S'/Y = -184(1/Y) + 0.22 + 0.0000082Y - 0.089(L/Y) + v',$$
$$\qquad\qquad (18)\ \ \ (0.020)\ (0.0000042)\ \ (0.0041)$$
$$R = 0.56 \qquad \text{est. } \sigma_{v'} = 0.33.$$

[10] F. N. David and J. Neyman, "Extension of the Markoff Theorem on Least Squares," *Statistical Research Memoirs*, Vol. II, Dept. of Statistics, University College, London ,1938, pp. 105–116.

Figures in parentheses below estimated parameters are their estimated standard errors. The reader can easily see from the large differences in estimates the importance of getting homoscedastic distributions before applying least squares methods to a set of observations. Entirely different conclusions about the effect of initial liquid assets on savings will result depending on whether we use (3.1.e) and (3.1′.e), or (3.2.e) and (3.2′.e). In the next section, the interpretation and significance of the differences will be more fully discussed.

The frequency distributions of S, Y, and L in the preceding section give the reader an immediate impression of the level and range of these three variables. In order to facilitate interpretation of (3.2.e) and (3.2′.e) we present means and standard deviations of the transformed variables $1/Y$, L/Y, S/Y and S'/Y.

	$1/Y$	L/Y	S/Y	S'/Y
mean	0.509×10^{-3}	0.526	0.205×10^{-1}	0.106
standard deviation	0.505×10^{-3}	1.815	0.347	0.395

The same types of equations have been estimated from the same sample using weighted instead of unweighted observations. There are two ways of interpreting the weighted calculations. The weights give each spending unit or stratum of units the same relative importance among sample observations as its relative importance among spending units or strata in the population. If the disturbances can be considered as drawings from a normal universe with zero mean and constant variance, a natural way of introducing the weights is in

$$p(v_1)^{w_1}p(v_2)^{w_2}\cdots p(v_N)^{w_N} = \left(\frac{1}{\sqrt{2\pi}\sigma_v}\right)^{\sum_1^N w_i} \exp\left(-\frac{1}{2\sigma_v^2}\sum_1^N w_i v_i^2\right);$$

i.e., as exponents of each term of the likelihood function. This interpretation applies to equations (3.2) and (3.2′). Maximum likelihood estimates of the parameters of the savings equation will then be obtained by minimizing $\sum w_i v_i^2$, with w_i = known weights. In terms of estimation equations this simply means that cross moments like $\sum w(S/Y)\ (1/Y)$ replace those like $\sum(S/Y)\ (1/Y)$ in the "normal equations" of standard least squares methods. The estimated variance of the disturbances becomes

$$\text{est. } \sigma_v^2 = \frac{\sum_1^N (\text{est. } v_i)^2 w_i}{\sum_1^N w_i}$$

instead of

$$\text{est. } \sigma_v^2 = \frac{1}{N} \sum_1^N (\text{est. } v_i)^2$$

in the conventional arrangement. As is well known, in linear regression equations the estimated parameter coefficients are linear functions of the observations of the random variable. In this formulation of weighted regressions, the estimated parameters are weighted sums of the observations; hence the sampling errors of the estimates will involve squares of the weights. In an equation of the form

$$x_i = \sum_{j=1}^{n} \alpha_j z_{ij} + v_i; \qquad i = 1, 2, \cdots, N;$$

with N sample observations carrying weights $w_1, w_2, \cdots, w_N$; we find

$$\text{est. } \sigma_{a_k}^2 = \text{est. } \sigma_v^2 \left[\sum_{j=1}^{n} (\Delta^{kj})^2 \sum_{i=1}^{N} w_i^2 z_{ij}^2 + 2 \sum_{j<m} \Delta^{kj} \Delta^{km} \sum_{i=1}^{N} w_i^2 z_{ij} z_{im} \right], \tag{3.3}$$

in which $a_k = \text{est. } \alpha_k$, and Δ^{kj} is the kj element of the inverse matrix of the moments,

$$\sum_{i=1}^{N} w_i z_{ij} z_{ik}.$$

Alternatively the same equations for estimating the parameters, but not their sampling errors, would be obtained if we were to minimize $\sum wu^2$ on the assumption that $\sqrt{w}u$ is a normal variable with zero mean and constant variance, $\sqrt{w}$ would have to be inversely proportional to Y in order that $\sqrt{w}u$ have uniform variance. Under this interpretation, the sampling errors of the estimated parameters, without correction for the loss of degrees of freedom in estimation of α_i, are given by

$$\text{est.} \sigma_{a_k}^2 = \text{est. } \sigma_u^2 \Delta^{kk} \frac{1}{N} \sum_1^N w_i, \tag{3.4}$$

Δ^{kk}, as before, is the kk element of the inverse matrix of weighted moments. Empirically, we have found practically no difference between (3.3) and (3.4). In the two variable case they can be shown to bear a very simple relation to each other such that appreciable differences are extremely unlikely.[11]

The foregoing remarks assume that we are dealing with a simple random sample. The Surveys of Consumer Finances are actually based on a more complicated multistage sampling scheme.[12] We have assumed that all the v_i are mutually independent but the nature of the sampling scheme, particularly the clustering, may cause some intercorrelation among the v_i. We have not allowed for such effects, but do not consider them to be of primary importance.

Oversampling, a main contribution to differential weighting as explained previously, should optimally be in proportion to the *standard deviation* of some variable whose mean it is desired to estimate as efficiently as possible. Oversampling in multiple purpose surveys like the Surveys of Consumer Finances is carried out in the light of heteroscedasticity in several variables and may not be completely adapted to our narrower objective of estimating patterns of consumer saving behavior. The variance of u in the savings function grows roughly as the square of income, with the standard deviation growing as income itself.[13] The actual weights fall linearly with income but not sharply

[11] The whole problem of the treatment of weighted regressions is one that has been taken up at length, in other contexts, in classical methods of least squares. In addition to the Neyman-David paper cited previously, representative discussion occur in the following references: W. E. Deming, *Statistical Adjustment of Data*, Wiley, 1943, Chs. II and IV; J. B. Scarborough, *Numerical Mathematical Analysis*, 2nd ed., The Johns Hopkins Press, Baltimore, 1950, Ch. XIV; E. W. Whittaker and G. Robinson, *Calculus of Observations*, 4th ed., Blackie, London, 1946, Ch. IX. Of course, nearly all least-squares methods are due, ultimately, to Gauss.

[12] See "Methods of the Survey of Consumer Finances," *Federal Reserve Bulletin*, July, 1950, pp. 795–809.

[13] Some indication of the relation of the variance of u' to income can be derived from the table below. The standard deviations given in the table refer to the nonentrepreneurial sample used throughout this study (1949 survey) combined with a similar group from the 1948 survey. Furthermore, they are standard deviations of the pooled residuals from seven different savings-income regression lines (for different asset and home-ownership groups), unweighted, using S'.

Disposable Income	Estimated Standard Deviation of u'
\$ 0– 999	\$ 275
1000–1999	324
2000–2999	487
3000–3999	680
4000–4999	854
5000–7499	1168
7500 and over	2288

enough to give a series inversely proportional to income. The ratio between the largest and smallest weight is roughly 5:1, while the ratio between the largest and smallest income level is much greater. For this reason it would be superior to regard wu instead of $\sqrt{w}u$ as a normal variable with zero mean and uniform variance. If the weights were chosen inversely proportional to income, and $u=vY$, then the minimization of $\sum w^2u^2$ would lead to proper estimation equations. Ideally, and without regard to cost, there should be much higher oversampling of the upper income groups. In practice, however, one can sample only some easily identifiable unit, such as a geographical point, and it is not possible to establish a perfect correspondence between place of residence and income.

Two criteria lie behind the use of weights in survey tabulations, bias and efficiency. Where possible it is instructive to distinguish between the two. In our problem, the possibility exists that the savings functions are not exactly the same in all strata. We interpret our equations as estimates of an average equation, an equation that depends on averages of the parameters of the different equations in the separate strata. The use of weights guards against bias in this case. If it were true that the equations were identically the same in all strata the problem of bias would not be important, but differential sampling rates with suitably chosen weights can in this situation be used to improve efficiency over unweighted calculations. When we use weighted observations for the estimation of (3.1) and (3.1′) we are trying to overcome both bias and inefficiency, but principally the latter. In equations (3.2) and (3.2′) the variables have already been transformed so that the variance of the disturbances is a constant and efficiency is not as great a problem as previously. Weighted estimates of (3.2) and (3.2′) are designed primarily to avoid bias.

Weighted estimates of equations (3.1)–(3.1′) and (3.2)–(3.2′), using only first powers of weights in the estimation equations, are

$$\text{(3.1.e.w)} \qquad S=-190+0.10Y+0.0000046Y^2+0.0050L+u,$$
$$\qquad\qquad (36)\ (0.016)\ (0.0000012)\qquad (0.0052)$$
$$R=0.36 \qquad \text{est. } \sigma_u=\$686$$
$$\text{(3.1′.e.w)} \qquad S'=-290+0.23Y+0.0000019Y^2+0.013L+u',$$
$$\qquad\qquad (34)\ (0.015)\ (0.0000012)\qquad (0.0049)$$
$$R=0.58 \qquad \text{est. } \sigma_{u'}=\$640$$

$$(3.2.e.w)\quad S/Y=-135(1/Y)+0.13\ +0.0000050Y-0.11(L/Y+v,$$
$$(15)\qquad (0.016)\ (0.0000037)\quad (0.0036)$$
$$R=0.71\qquad \text{est. } \sigma_v=0.27$$
$$(3.2'.e.w.)\ S'/Y=-164(1/Y)+0.21\ +0.0000104Y-0.11(L/Y)+v',$$
$$(19)\qquad (0.020)\ (0.0000046)\quad (0.0045)$$
$$R=0.64\qquad \text{est. } \sigma_{v'}=0.33.$$

These equations are very close to the unweighted equations although there are some noticeable differences. For many purposes the use of weights will not make much difference, but in the most refined attempts to get an unbiased estimate of the savings function for the economy as a whole, weighted data are recommended. A major effect of weights is to reduce the estimates of σ_u and σ_u'. In general, univariate distributions and their associated moments are affected by weighting.

If we proceed on the assumption that wu instead of $\sqrt{w}u$ is a normally distributed random variable with uniform variance, we estimate (3.1) and (3.1′) by minimizing $\sum w^2u^2$. The results are

$$(3.1.e.w^2)\qquad S=-148+0.076Y+0.0000058Y^2+0.0030L+u,$$
$$(34)\ (0.016)\quad (0.0000015)\qquad (0.0059)$$
$$R=0.31\qquad \text{est. } \sigma_u=\$635$$
$$(3.1'.e.w^2)\quad S'=-263+0.22Y+0.0000033Y^2+0.015L+u',$$
$$(32)\ (0.015)\ (0.0000015)\qquad (0.0056)$$
$$R=0.55\qquad \text{est. } \sigma_{u'}=\$607.$$

These results do not differ greatly from those obtained in (3.1.e.w) and (3.1′.e.w); furthermore we find that they differ from the estimates of (3.2) and (3.2′) in the same way that (3.1.e.w) and (3.1′.e.w) do. This conforms to our feeling that the oversampling and subsequent weighting of observations in the Surveys of Consumer Finances is not sharp enough to eliminate the heteroscedasticity in weighted data on savings.

A major difference between the estimates of (3.1)–(3.1′) and (3.2)–(3.2′) is in the marginal relationship between L and S. The former estimates imply that this relationship is either insignificant or slightly positive, while the latter imply that it is significantly negative. The equations (3.2.e)–(3.2′.e) or (3.2.e.w)–(3.2′.e.w) give a more correct picture of consumer behavior because these estimates are derived by a

statistical method which is more nearly consistent with its underlying assumptions than is the case for the other estimates. Yet there is a reason for the differences in economic behavior. People in low income groups are more likely to be influenced in their savings decisions by their asset holdings than are people in the upper groups. Indeed, the asset effect could even change direction between low and high income groups. The estimates of (3.2.e)–(3.2′.e) and (3.2.e.w)–(3.2′.e.w) allow the low-income observations to play a strong role in the calculated form of the equations, while the outlying observations in high income groups dominate (3.1.e)–(3.1′.e), (3.1.e.w)–(3.1′.e.w), and (3.1.e.w^2)–(3.1′.e.w^2). The weights are not sufficiently variable to diminish the extreme contribution of the upper income groups to the regressions in (3.1.e.w)–(3.1′.e.w) or (3.1.e.w^2)–(3.1′.e.w^2). An additional comment may be in order. Since the upper income groups make the greatest contribution to saving—the top decile of the income distribution having done 105 per cent of the net saving in 1949—an optimum estimate of the structural relation between saving, on the one hand, and income, assets, etc. on the other, may not be more efficient in predicting aggregate savings than a nonoptimal method which allows the high income recipients to dominate the relation. The use of nonoptimal methods for obtaining a relation between saving and other variables is mentioned only because economists frequently wish to estimate aggregate saving on the basis of information only about aggregate (or mean) income, assets, etc. When the optimally estimated structural relation is nonlinear, involving, for example, an income-asset (interaction) term, the joint distribution of income and assets is needed in order to derive estimates of saving. While such information was available for the (past) years which were used in estimating the structural relations, it will probably not be available, or easily predictable, for future years for which saving is to be estimated. In this case, a nonoptimal method which allows the high-income, high-asset people who dominate aggregate saving to dominate the estimates of the parameters may give an equation which predicts aggregate saving more closely than an equation based on better methods.

We are naturally led to question whether we would get more similar results between the two types of equations (3.1)–(3.1′) and (3.2)–(3.2′) if we modified them to permit the marginal asset effect to vary with the income level as in (1.2). Our alternatives are thus

$$S = \beta_0 + \beta_1 Y + \beta_2 Y^2 + \beta_3 L + \beta_4 LY + u, \tag{3.5}$$

$$S' = \beta_0' + \beta_1' Y + \beta_2' Y^2 + \beta_3' L + \beta_1' LY + u', \tag{3.5′}$$

and

$$(3.6) \qquad S/Y = \beta_0(1/Y) + \beta_1 + \beta_2 Y + \beta_3(L/Y) + \beta_4 L + v,$$

$$(3.6') \qquad S'/Y = \beta_0'(1/Y) + \beta_1' + \beta_1' Y + \beta_3'(L/Y) + \beta_4' L + v'.$$

The estimated equations from the same sample are (using unweighted data):

$$(3.5.e) \qquad S = -247 + 0.14Y - 0.00000065Y^2 - 0.019L + 0.0000052LY + u,$$
$$\qquad\qquad (38)\ (0.016)\ (0.00000153)\ (0.0085)\ (0.0000015)$$
$$R = 0.49 \quad \text{est. } \sigma_u = \$721$$

$$(3.5'.e) \qquad S' = -319 + 0.25Y - 0.00000006Y^2 + 0.0048L + 0.0000011LY + u',$$
$$\qquad\qquad (36)\ (0.016)\ (0.00000147)\ (0.0082)\ (0.0000015)$$
$$R = 0.61 \quad \text{est. } \sigma_{u'} = \$693$$

$$(3.6.e) \qquad S/Y = -105(1/Y) + 0.13 - 0.000011Y - 0.13(L/Y) + 0.000034L + v,$$
$$\qquad\qquad (14)\ (0.016)\ (0.0000033)\ (0.0039)\ (0.0000019)$$
$$R = 0.67 \quad \text{est. } \sigma_v = 0.26$$

$$(3.6'.e) \qquad S'/Y = -135(1/Y) + 0.22 - 0.0000080Y - 0.13(L/Y) + 0.000035L + v$$
$$\qquad\qquad (17)\ (0.019)\ (0.0000041)\ (0.0048)\ (0.0000023)$$
$$R = 0.61 \quad \text{est. } \sigma_{v'} = 0.31.$$

The introduction of a possible interaction effect between L and Y does not alter our basic conclusion that the results depend heavily on the choice of statistical method. The assumption of homoscedasticity when it is, in fact, not true appears to give misleading results for at least some parameters.

When the variable LY is added to the savings equation the coefficient of Y^2 changes from positive to negative, which is consistent with our equation (1.5) derived from a combination of (1.3) and (1.4). If we break up the derived equation into its components we see that the estimated coefficient of Y^2 in (1.3) has the same absolute value as in (1.5) but with opposite sign. In (3.6.e) and (3.6'.e) the estimates of the parameter in question are both significant, although this is not true in (3.5.e) and (3.5'.e). There is a problem of intercorrelation involved. LY, and Y^2 are correlated, as are L and Y; therefore we are left with some doubt about the relative contributions of the separate variables in the estimated equations.

On dropping Y from (3.6.e)–(3.6'.e), we get

$$(3.6^*.e) \qquad S/Y = -83(1/Y) + 0.084 - 0.13(L/Y) + 0.000032L + v,$$
$$\qquad\qquad (12)\ (0.009)\ (0.0039)\ (0.0000018)$$
$$R = 0.67 \quad \text{est. } \sigma_v = 0.26$$

$$(3.6^{**}.e)\quad S'/Y = -119(1/Y) + 0.19 \quad -0.13(L/Y) + 0.000033L + v'.$$
$$(15) \qquad (0.010) \quad (0.0048) \qquad (0.0000022)$$
$$R = 0.61 \quad \text{est. } \sigma_{v'} = 0.31.$$

In view of the close general agreement between weighted and unweighted calculations we have not made estimates of the equations with the LY variable in the weighted case.

4. DISCUSSION OF FINDINGS

In view of the large differences obtained from the alternative estimation procedures, we find it useful to summarize other calculations we have made based on different methods, a different form of the savings equation, or different data.[14] An alternative method is to compute savings-income regressions by asset classes or savings-asset regressions by income classes. The former are done for data referring to 1947, 1948, and 1949; the latter for data referring to 1948 alone. The computed regressions show an increasing savings-income slope as assets increase and an increasing savings-asset slope as incomes increase. This checks the positive coefficient of LY in (3.5.e)–(3.5′.e). The general magnitudes of the savings-income and savings-asset slopes agree with the estimates in our equations.

From a reinterview sample of 606 urban spending units, a subsample in the 1949 Survey of Consumer Finances, we find estimated coefficients of L/Y and L roughly the same as in (3.6.e). The reinterview sample, however, uses contemporaneous response from the first interview rather than memory response from the second interview for the variable L, and yields coefficents slightly larger in absolute value than those of (3.6.e). The reinterview calculations are made for a different form of savings equation, namely, one that contains additional variables such as age, household size, income change and home-ownership status. It also uses log Y instead of $1/Y$ and Y. The savings-income slope estimated from this equation agrees with that in (3.6.e). An equation calculated from the 1950 Survey of Consumer Finances similar to that from the reinterview sample again confirms our estimated coefficients in (3.6.e). In both these two other samples we find that the residual variation from our equations has approximately constant standard deviations in all income deciles except the lowest, where a few extreme cases make for larger variability.

Our findings are intended to have both methodological and substan-

[14] See L. R. Klein, *op. cit.* and J. N. Morgan, *op. cit.*

tive content. The methodological points concern the estimation of parameters in equations with heteroscedastic disturbances and the use of weighted observations. In the theoretical literature there has never been any question about appropriate estimation methods to be used when ratios among variances are known. The need for uniformity of variance or appropriate correction for nonuniformity has not, however, always been appreciated in empirical work. Evidence of this is the widespread calculation of regressions of savings or expenditures on income and other variables in cross-section studies, ignoring the effects of heteroscedasticity. We find in data from the Surveys of Consumer Finances, that minimization of $\sum u^2$ leads to quite different results, for asset coefficients, from minimization of $\sum v^2$ where $\sigma_u^2 = Y^2\sigma_v^2$.

The weights used in our sample do not have much effect on estimates of the savings equation in either the homoscedastic or the heteroscedastic cases. The weights do not vary enough to account for the change in variance of disturbances between high and low income or asset groups; moreover, the parameters of the savings equation do not vary greatly among weight groups. Insofar as differential sampling rates are used in the Surveys of Consumer Finances to improve statistical efficiency, our findings suggest that the differentials are too mild for estimates of statistics on savings. We do not, however, cast any light on the efficiency of the sample design for estimation of statistics on other variables.

Our main substantive findings relate to the estimation of parameters of savings behavior. We are particularly interested in the nonlinearities and asset effects in savings equations. One nonlinearity of importance is the curvature in the savings-income relation, other things constant. Equations (3.2.e)–(3.2.e.w) and (3.2′.e)–(3.2′.e.w) all show positive curvature, although some of the estimates have large sampling error. If we interpret (3.6.e)–(3.6′.e) as estimating the model in (1.3)–(1.5), the estimates of the positive curvature are more significant. One reason for interest in these particular parameters is that they are of use in assessing the effects of income redistribution on saving.

The regressions in the heteroscedastic cases show a positive or insignificant net relation between L and S or S', contrary to much current economic theorizing. By switching to a form of the savings equation which has homoscedastic disturbances, we are able to specify the asset effects more definitely. The other type of nonlinearity studied, namely, the interaction effect between L and Y on S or S', is interesting in itself as a nonlinearity but also as an indicator that the much discussed

relation between wealth and savings is not the same for all persons. The negative value of the asset effect falls as income rises and reaches zero near $4500 disposable income, at which point it becomes positive. This is a quite reasonable result, but has not, to our knowledge, been adequately accounted for in economic analysis. Although the Survey measure of liquid assets does not encompass total individual wealth, we should like to point out that there is a strong correlation between net worth and liquid assets of individual households. We do not use a net worth variable because it is less reliably estimated than liquid assets and only in broad classes. The exclusion of farmers and businessmen from our sample automatically excludes those households owning the largest nonliquid assets.

The various alternative calculations made in this paper do not enable us to select either S or S' as a superior variable for the study of savings behavior. A separate study of durable goods expenditure does reveal two respects in which this type of purchasing behavior differs from savings behavior. The asset effect on durable goods expenditures is opposite in sign to that on savings. At very low income levels, say, near zero, expenditures on durables fall to zero, while savings are substantially negative. Behavioral differences like these are obscured by the use of S', but even the components of S are not all homogeneous, and other definitions of saving may prove to be more fruitful. Nonetheless, regardless of the resolution of the problem of defining savings concepts we do feel that our calculations throw some light on the other problems we initially posed.

AN ECONOMIC APPRAISAL OF PRICE MEASURES*

RICHARD B. HEFLEBOWER
Northwestern University

THE significant and varied roles which measures of price changes have come to play in the analysis and administration of the modern economy require that they be adapted to playing those parts. That is, of course, a maxim for those who prepare the price measures. But what is argued here is addressed equally to those who use price measures. One might say with justice that such an appraisal is made necessary by the fact that interest in price measures is no longer limited to the specialist. Price facts now provide part of the background information, and even formulas, which enter into public and private decisions. Millions of dollars in private commodity or wage contracts turn on movements of published price series or indexes. Congressional hearings abound with price movement comparisons such as between farm and retail prices or the latter and wages. Equally important, however, are questions about economists' practice, for they now use price measures for such diverse purposes as putting dollar figures in real terms, comparing movements of prices among vertical steps in the industrial structure, or studying the timing of the response of different types of industries to dynamic influences. Altogether price measures are used for a variety of purposes to indicate what goes on in a variety of markets in response to a variety of dynamic influences.

Yet whether the reference is to a price series or to a composite index number, a single type of measure, conceptually, is all that is ordinarily available. This is the reported price for a specified item frequently used to represent a product line, or else an index number for a set of such prices, that is, the weighted average of the reported prices for a representative, fixed bill of carefully specified goods. It is true that some widely-used indexes such as that of hourly earnings and that of prices paid by farmers depart from this standard, but to that extent they are considered to be defective.[1] Whether or not it conforms to this objec-

* Presented at the 110th Annual Meeting of the American Statistical Association, Chicago, December 1950.

[1] Such "defects" usually reflect the shortage of funds now or at an earlier date for obtaining price data for commodities or services which conform to specifications. Thus the Bureau of Labor Statistics' only current and comprehensive measure of wages is computed by dividing total payrolls reported by the corresponding manhours reported. Consequently the number of employees in each occupation in each reporting period becomes the weight for averaging hourly earnings for that period. The Bureau's Index of Urban Wage Rates, which conforms to the standard described here, has been prepared for only a limited job and time-period coverage. This is based on "occupational wage" surveys. See the Bureau's *Bulletin* No. 993, *Techniques of Preparing Major B.L.S. Statistical Series*, 1950, pp. 37–41 and 64–72.

The "Index of Prices Paid by Farmers" computed by the Bureau of Agricultural Economics obtains price reports from merchants on "size, grade, and quality more commonly sold to farmers" and is influenced by changes in what farmers "commonly" buy. (U.S. Dept. of Agric. Miscellaneous Publication No. 703, p. 121.) See also O. V. Wells' paper read before this Association *on December* 30, 1949.

tive a single type of measure is generally used for and often descr ed or implied as applicable to all problems.[2] Much of what is said he is concerned with the "all purpose" usability of any price measure.

This emphasis on the objectives for which measures are used and character of the phenomena they record is the theme of this paper. present writer has little to add on the subject of index number form las *within the context* in which they are ordinarily discussed. Nevertheless in Section I the character of these measures and some practices in their application are explored to make clear what they do and do not measure. Conclusions on these points prove useful in Sections II and III. In Section II the variety of uses of price measures is outlined and related to the development of economic ideas and policies. Section III presents a sketch of the varied forms of market adaptation to dynamic influences, from which emerges conclusions as to price changes which are equivalent to a movement of prices for certain purposes, but which are not portrayed by the usual measures. With the preceding parts as background, some suggestions for price-measure improvement are presented in Section IV, and in Section V comments are made on appropriate use of some existing indexes.

I. ATTRIBUTES OF PRICE MEASURES

The accepted concept of price and the measures based upon it involve a distinction between a value change and a product or output change. Price measures are designed to record the changes in the number of dollars exchanged for a carefully specified commodity or a fixed bill of goods. In the simpler case where only one price is considered, "*reference to a 'price' must always rigorously designate the corresponding 'product'* if clarity is to be maintained."[3] A corresponding index number is the one constructed by use of common weights[4] and a fixed bill of such products or services for the two or more time periods being compared.

These requirements are imposed on price measures so that they will record "pure price" changes. As far as such measures are concerned, all developments which affect aggregate dollars spent or received but which do not reflect more or less dollars exchanged for the specified

---

[2] The reference here is to the formal structure of the measure, not to the items or commodities included or to the vertical stage at which prices for particular commodities are used.

[3] Commitee on Price Determination of the Conference on Price Research, *Cost Behavior and Price Policy* (National Bureau of Economic Research, 1943), p. 39.

[4] The term "common weight" will be used here generally in preference to "fixed weight" for the latter term is used widely to refer to indexes whose weights remain unchanged for an extended period. Here, however, some of the questions raised apply equally in principle although not in degree to link indexes in which weights may be changed each year but are "common" for both periods in each link.

bill of goods are relegated to the output category.[5] Consequently, Fisher's "ideal" formula[6] for prices is rejected because it reflects not merely movements of prices of products but also their changing importance.[7] Instead, the accepted doctrine is that "Major changes in the specifications of commodities, shifts in the relative importance of sales to different types of purchasers or by different types of sellers, alterations in the distribution pattern of the industry, or changes in commodities with economic conditions are handled in each instance by 'linking' so as to prevent movement in the index for the period in which the change is made."[8]

In the application of these standards for price measurement two types of questions arise. One has to do with the effect on those measures of the exigencies of data collection and analysis. The other type of query has to do with whether dynamic developments which accompany price movements distort what price measures, as actually constructed, show.

1. *Reported vs transaction prices.* Most of the price data available are those which interested persons, usually sellers, are willing to report. Does the relationship between reported and transaction prices differ so much among commodities, or for a given commodity at different times, that price movements can be misinterpreted substantially?

2. *Specification of items.* The exigencies of dealing with product specifications[9] and changes therein have led to two practices.

a. Cognizance is taken only of significant specification changes.[10] Does this introduce bias in those commodity areas in which product changes take place by a series of small developments, which often alter cost and utility? In such cases, prices as measured would be different, usually higher, than they should be if pure price change were accurately recorded.

---

[5] Indeed this is what is done when changes in aggregate dollars, such as of national income, are divided by a price index to obtain "real" income. But as Irving Fisher pointed out (*The Making of Index Numbers*, 1922; p. 79 ff.), no common weight quantity index would include all of the residue of an aggregate value change not reflected by the price index. The price index times the quantity index would not yield the aggregate value index precisely.

[6] *Ibid.*, p. 220 ff.

[7] W. M. Persons, "Fisher's Formula for Index Numbers," *Review of Economic Statistics*, Preliminary Volume III, 1921, p. 12, and W. C. Mitchell, *Index Numbers of Wholesale Prices in the United States and Foreign Countries* (Bulletin 284, U. S. Bureau of Labor Statistics), p. 92, 93.

[8] B.L.S. *Bulletin* No. 993, p. 28. In the same reference the details of this "linking" procedure are explained.

[9] The stress is on physical specifications for sale to a particular class of customers on particular terms. That brands are not considered important is shown by the practice in the Consumers Price Index of permitting the level of price of a new brand replacing an old one for an item of the same physical quality, to be reflected in the index (B.L.S. *Bulletin* 993, p. 2). There is considerable variation in the rigor of physical specifications. Thus wheat is quoted by grade but not by protein content, but steel sheet specifications are given in detail.

[10] B.L.S. *Bulletin* No. 993, p. 28.

b. Prices are collected only for the more standardized items of such complex, constantly-changing, product lines as machinery. Are not non-standard goods often more important and more or less volatile in price in some fields? Does this practice also mean that product coverage and reliability of price measures is progressively less the more highly processed is the commodity?

3. *Representativeness of items priced.* Whether the reference is to a single product in the physical sense or to a composite index, price data are obtained for fewer "items" than the area they are to represent.[11] With but few exceptions a given product is sold in two or more physical variations, and certainly on more than one set of sales terms. Selecting representative items for a given period is a subject on which nothing is added here. Rather the concern is with whether the very conditions which bring about a price change also work to make the items priced less representative of the area they are presumed to portray.

4. *Fixed weights.* Similarly the question arises as to whether the time of price change is also one of such divergencies in price movements and in quantity movements for various items in a product line or in a commodity field that weights of an earlier period give erroneous results.[12]

Running through these queries about the application of price measurement theory has been a common thread concerning the possible effect of dynamic developments associated with price change on the reliability of price measures. Indeed the disturbing thought arises that measures of prices may be least accurate when they are needed most; that is, at the time of actual, if not always measured, price change. In part this would stem from inadequacies of price data as under (1) above. The remainder involves the ideas of fixed representativeness or fixed weights. In the end we shall find that this raises doubts about the appropriateness of a "pure price" measure for some situations. The importance of all of these questions depends in part on what indexes are to be used for. In addition the kind of dynamic adjustments made in the areas where prices are to be gauged will be found to be relevant.

---

[11] Consequently, even though most of the points made in this paper are directed to index numbers, they apply in nearly all cases to individual price series.

[12] It should be emphasized that a change in importance of products *per se* does not introduce error in an index; it is only when products of increased significance differ in price movement from those of declining importance that out-of-date weights affect the current use of a "pure price" index.

II. USES OF PRICE MEASURES AND THEIR RELATION TO ECONOMIC IDEAS

The second step is to outline the varied uses to which price measures are put, and to show how these are related to the evolution of economic doctrine. Measures of price change are used:

1. To gauge the effect of price movements on changes in welfare between adjacent periods, or between dates quite distant from each other, such as (a) when gross national product, national income, or income of any group is "deflated" to be put in "real" terms; (b) when "basic wage rates", or rates for specific jobs, are to be adjusted for the "cost of living" under a union contract or government mandate; (c) when relative prices are used to indicate relative incomes, such as in the farm "parity index"; or (d) when price indexes, or individual price series, are used to revise prices in contracts.

2. To indicate when there has occurred a change in the relationship of demand and supply, in the study of such problems as (a) the turning point in cyclical movements of the economy as a whole, for which indexes of "sensitive prices" have sometimes been used; (b) the turning points at various vertical steps in the production and distribution in the economy;[13] or (c) the turning points for various commodity areas viewed at the same vertical stage, such as food versus metal products.

3. To show the promptness and extent to which adjustments to demand and cost movements by various industries take the form of changes in prices[14] (a) vertically among stages from raw materials to retail sales. such as in the "marketing margin" studies of the Bureau of Agricultural Economics;[15] (b) horizontally among industries in the same vertical step, such as finished-goods manufacture, such as during the "price rigidity" controversy of the early 1930's.

4. To portray the price record of an industry as part of the analysis of its contribution to economic progress, the theme which Nourse has emphasized.[16]

5. To study the extent to which price changes explain the symmetry or asymmetry in the movement, usually between adjacent periods,

[13] Such studies were of a very practical value in 1941–42 when the Office of Price Administration was considering the effect of simultaneous freezing of prices at all vertical steps.

[14] Here the emphasis should not be on product prices themselves, but on their movement relative to some other variable such as factor prices as A. C. Neal has pointed out (*Industrial Concentration and Price Inflexibility*, American Council on Public Affairs, 1942, Chapters III and IV).

[15] Price indexes were used to judge a lag in vertical adjustments in *The Economic Report of the President*, January 1949, p. 6.

[16] E. G. Nourse, *Price-Making in a Democracy* (Brookings, 1944).

of such value aggregates, as (a) consumer incomes and retail sales as in 1947 when there was wide concern about the adequacy of consumer income; (b) labor's income and value of consumer goods sold; or (c) farm income and value of goods sold to farmers.

Interest centered on the first two of these uses at the time that the inflation of World War I gave the first great stimulus to quantitative price research. Price indexes deflated dollar values to get a measure of welfare (1 a of above classification of uses of price measures), and recorded price change as an indicator of the supply-demand situation (2 a, primarily). The indexes which were developed at that time and largely adhered to since, reflected these objectives and were in harmony with the view then taken of economic processes. The received doctrine emphasized the tendency toward a beneficent equilibrium brought about by a series of particular market adjustments, in which price is the means of establishing balance. Furthermore since nearly all markets were viewed as competititve, in a loose meaning of that term, price movements, not output or product-type changes, were viewed as the chief evidence of a change in supply or demand. Movements of prices for the economy as a whole were held to be monetary in origin, even when these prices changes were of the short term cyclical character. These are all phases of the competitive, full employment economics of that era.

In the past two decades, however, our view of the economy and of particular markets has come to emphasize disequilibrating tendencies. For the economy as a whole the business cycle, or the socially disrupting "underemployment equilibrium" or inflation, overshadows equilibrium at full employment in both analysis and policy. As a means of analysis emphasis has shifted from disequilibria among industries, or between money supply and output, to lack of balance between such value aggregates as consumers' income and value of output.[17] In such doctrine, too high commodity prices or too low wage rates, as one determinant of the respective aggregates (use 5 above) could be the origin of inadequate consumption or investment incentive. Equally important is the recognition of the pervasiveness of "imperfect competition" which centered attention on the price change record of such undustries as a source of unbalance in the economy[18] and as an indicator of the

[17] See, for example, *Economic Report of the President*, January 8, 1947, pp. 10–11, and July 21, 1947, p. 31.

[18] Particularly emphasized in the "price rigidity" controversy set off by G. C. Means' *Industrial Prices and Their Relative Inflexibility* (Senate *Document* No. 13, 1st Session, 74th Congress, 1935). For a critical review of the literature on this subject see Neal, *Op. cit.*, chapter II.

social performance of particular industries or commodity areas[19] (use 4). These new emphases in the use of measures of price change differ from the older (uses 1 and 2) in two regards: Short term and often finely-measured movements are stressed. Consequently measures which were once satisfactory even though they were rough approximations and belatedly reflected price movements have become inadequate. Second, the possibility arises that objectives 3, 4, and 5 cannot always be met even if the data are good, because of the concurrent shifts in what is sold and on what terms. For that reason relevant aspects of the structure of markets and prices and of methods of adaptation to dynamic influences must be explored.

III. CHANGES NOT MEASURED BY CONVENTIONAL DEVICES

A basic feature of most markets, except those for such raw materials as farm products, is that the market for a particular commodity is broken into segments which over time have a fluctuating price and quantity relationship to each other. What is sold in the various segments may be physically identical and distinguished only by brand or package, but often there are also quality, appearance, or functional differences. This segmentation represents the net effects on market structure of diversity in final buyer wants, including sometimes the contrast between industrial and final consumers as purchasers, of variety in distributional arrangements, and of manipulation of both of these by the sellers; and the whole structure is rarely static.

The cost-price margins vary among these market segments because of differences in the price-making forces. Quite generally the dollar margin and often the percentage margin between direct costs and price, and between total cost and price, correlates positively with the height of the selling price in the market segment. The extent to which this characterizes American industry was revealed, when under conditions of rising direct costs and demand, the Office of Price Administration froze prices or margins. Both the consequent shift to the more profitable segments and the accounting investigations of the agency showed how generally margins differed among segments, and how margins correlated with prices.[20] Some of these margin differences are temporary.

[19] T. J. Kreps, *The Social Performance of Business* (TNEC *Monograph* No. 7) and S. Nelson and W. G. Keim, *Price Behavior and Business Policy* (TNEC *Monograph* No. 1) Chapter II.

[20] Some of the more significant industry examples are summarized in *Problems in Price Control: Changing Production Patterns*, edited by Benjamin Caplan (Office of Temporary Controls, published by U. S. Government Printing Office, 1947). See particularly pp. 10–12, 107–9, 128–36, 183–92, and 295–92.

Additional evidence is contained in *OPA Economic Data Series* (Office of Temporary Controls, Published by U. S. Government Printing Office, 1947), No. 5, Table 2; No. 13, Tables 4 and 5; No. 14, Table 4; and No. 25, Tables 2 and 7B.

More of them are of the persistent sort and reflect, among other things, the motivations and competence of the buyers, the role of various types of distributive trades, the success in differentiation, and the time lag in the spread of buyer information. This is another way of saying that this diversity of margins exists because of differences in the price-forming factors or in sellers' price policies, whichever explanation one prefers.

For our purposes the significance of these structural aspects of markets and prices lies in what happens to them under dynamic conditions. Both the price-cost and quantity relationships among market segments are subject to two kinds of changes. One we shall term "structural" which stems from changes in the economic character of the various market segments. A potent force of this source is the spread of buyer information which leads them to conclude that what is offered in a lower-priced market segment is satisfactory. Significant also in the last thirty years have been distributional developments, which have expedited the process of consumer shift to low-priced market segments in part by introducing oligopsony in dealings with manufacturers, but more significantly by introducing retailer-brands. Regardless of the origin, these structural changes cause a realignment of relative quantities and often of relative prices, too, among the various market segments. Consequently fundamental aspects of an industry's performance either in the primary[21] market or at retail cannot be traced by the price in one market segment such as one price line or one type of brand distributed through one channel of sale.

Even though the ways in which various industries react to short-term movements of factor prices and of demand differ widely the present writer's investigations lead him to the following generalizations: The role of a varying relationship between quoted and transaction prices is probably less in industries of frequent than of infrequent changes in quoted prices. Frequency of quoted price change in turn is most clearly characteristic of industries having a narrow manufacturing margin over fluctuating raw material prices which tends to mean the industries processing farm products. Likewise those market segments of a product line in which prices yield only a narrow margin over

[21] Between 1926 and 1941 the percentage of rubber tires sold under distributive brands, such as those of chain stores, rose from 9 to 31 per cent of all replacement sales. (Warren Leigh, "Automotive Tire Sales by Distribution Channels," Study 50, University of Akron, Bureau of Business Research.) During the earlier part of this period tires sold under distributive brands were priced at about 25 per cent below those carrying manufacturers' brands and sold through independent dealers, but by 1947 this percentage was nearer 15 according to investigations by the present writer. Yet for this whole period the published wholesale price series for tires was based entirely on the latter brands and channel of sale.

direct costs are more apt to experience overt price movements when those costs change than are the wide-margin segments. In the latter, or even in all segments, price concessions are more apt to be used when demand declines without an equivalent drop in direct costs than when the latter are falling sharply also. But if the demand decline persists, quoted prices will drop tardily to the level of transaction prices.[22]

Then there are areas where the means of adjustment is chiefly in the products rather than in prices directly. In "price-lined" goods, this is the only way to make small cost-price adjustments.[23] And in most non-food consumer goods, such as apparel and furniture, or producers' goods such as castings or machinery, where physical characteristics are in constant flux, the amplification of this process to reflect minor factor-price changes, and even major demand movements, can be expected. These cases raise no conceptual problems. But the administrative task of keeping abreast of product modification is so great that my guess would be that current indexes reflect less price change than occurs in such commodity areas. Or where in desperation prices are collected only on a few standardized items from a product line, the prices of which are highly competitive and which happen also to provide only narrow margins over direct factor costs, the chances are that price movements for the line are over-reported.

Both in its impact on the economy and in the conceptual issues raised, this problem is minor compared to what we now know are sizeable shifts in "product-mix" and "sales-mix". These terms, popularized by the joint OPA-WPB effort to deal with them during the last war, refer to changes in the relative proportion of volume represented by various items in a product line, or sold to different classes of customers. The phenomenal mix shifts of the war period were only an amplified form of what goes on in peace time when there occur short-term changes in demand and costs in particular industries or in the

---

[22] This cannot be tested by comparison of "census prices" (explained on p. 474 below) and Bureau of Labor Statistics Prices. The former are so much affected by shifts in the "product mix" and "sales mix" (see p. 469 below) that the disparity between their movement and B.L.S. reported prices cannot be assigned to varying amounts of price concessions. The present writer has carefully checked rubber tires, margarine, and rolled oats prices and found a distinct tendency for price concessions of 1930 and 1931 to become overt price cuts by 1932 and 1933. The report prepared by the Bureau of Labor Statistics, "Consumers' Prices of Steel Products," published in the April 24, 1946, issue of *The Iron Age*, shows how steel price concessions of 1939 were wiped out in the defense boom of 1941.

[23] Where price-lines are rigidly adhered to, as in some items of apparel and furniture, a price change by changing the number of dollars to be charged for an identical item, requires that the item be moved to the next price line. This would be a greater change than most cost movements warrant; small cost movements can, therefore, be adapted to only by changing the product.

economy.[24] Indeed the shift in "mix" is often on the initiative of the buyers rather than of the sellers.

Regardless of origin such mix changes have a powerful influence on sellers and buyers. Through this process sellers' realization relative to costs and relative to physical inputs can move radically without a change in the transaction price of any item in the line. This is because of the varied margin over the different elements of cost which the respective items or market segments yield.[25] If volume shifts toward higher margin items or channels, for example, either profits will rise or an increase of factor cost can be absorbed. Simultaneously the amount of money buyers spend during a given time period to get what satisfies their wants may differ by a change in their choice among close substitutes rather than because of a movement in the prices of any one of these items.

What I have sketched here are the impure, imperfect markets we have and how they operate under dynamic conditions. What is shown contrasts with competitive markets, or if not that, a static pattern of imperfect markets, which are assumed in most price measures. Taking all three of the problems outlined here together—quoted vs transaction prices, product modification, and "mix" shifts—we reach two disconcerting conclusions. The first has already been anticipated; our price measures are apt to fail us when we need them most, that is when there is a movement of demand and costs. It is on these occasions that the developments outlined here, but which are not protrayed by conventional measures, are accentuated. Second, comparisons among industries, and among vertical stages in some commodity areas by conventional devices are apt to be so vitiated by the developments stressed here that those measures may give more misinformation than fact.

Whether what price measures show is erroneous depends on the pur-

---

[24] Many examples of this could be cited. The proportion of premium to all gasoline sales rose from 10 in 1940 to 33 in 1948, but fell back to about 22 per cent in 1950, as is shown by monthly reports of the Bureau of Mines. The percentage of replacement tires sales, represented by "first" line and premium, rose from 51 in 1940 to 58 in 1941, while the percentage of lower qualities fell correspondingly according to a survey by the Office of Price Administration now in the files of the National Archives. The percentage of all washing machines, which sold for over $70.00 at retail, fell from 37 in 1937 to 27 in 1939, but then rose from 29 per cent in 1940 to 43 in 1941 as is shown in the annual survey issues of *Electrical Merchandising* (McGraw-Hill.)

[25] For example, prime costs (direct labor and materials) represented typically about 70 per cent of the selling prices of women's slips in the $5.00 to $9.99 per dozen range, 67 per cent for the $10.00 to $14.99 range, and 63 per cent for the $15.00 to $19.99 group. Dollar margin differences, which may be more relevant, were proportionally greater. (See Office of Temporary Controls, OPA, *Economic Data Series No. 13, Table 4.*) Or to use the shift in sales pattern of washing machines, referred to in the preceding note, this development alone would enlarge the average dollar margin per machine between factory cost and selling price by over a quarter, according to cost information the present writer summarized from Office of Price Administration records.

pose for which they are to be used and on the particular markets and current developments with respect thereto. The significance of purpose needs to be underscored. For purposes which require a measure of "pure price" change, conventional measures are most apt to err when transaction prices move differently than do reported prices.[26] In contrast, where the objective is to obtain a valid understanding of what happens under dynamic conditions at the different vertical stages in a commodity area or among various industries at the same vertical stage, conventional indexes, even when constructed from transaction prices, are apt to give a distorted report. The reason is that the forms of market adaptation outlined above are highly significant at later processing stages while "pure" price changes are more usual in the raw materials markets. Similarly "pure" price change whether measured by reported or transaction prices—typifies some industries such as the food processing much better than others such as metals manufacturing. Then there are the broader-scope analyses in which the interest in prices is really in the price component of such aggregate dollar figures as incomes or value of output or of retail sales. From what we have emphasized about price indexes and about commodity markets it is clear that the movment of such dollar figures would differ substantially from what they would be estimated to be by multiplying production indexes by available price indexes. This last point, particularly, will be clarified in connection with some suggestions for improving price measures.

IV. SOME SUGGESTIONS FOR PRICE-MEASURE IMPROVEMENT

Both the administrative and conceptual improvement of price measures are difficult. The suggestions I offer are far from complete, and the last one to be made will give rise to sharp debate, I am certain.

(1) The time is past when we can view with complacency the disparity between transaction and reported prices for specific commodities. As one writer has put it, "economists using published price series are often the jest of industrial specialists."[27] My observation is that businessmen nearly always have at hand better price facts than those published about their commodities. This is particularly true when demand or cost movements presage a price change, the very time when price measurement is most significant. Yet attention-getting research on the differential price performance of sectors of the economy has depended almost exclusively on reported prices.

---

[26] See p. 472 below for the effect of weight changes on short-term "pure price" comparisons.

[27] A. R. Burns, "The Organization of Industry and the Theory of Prices," reprinted in *Readings in the Social Control of Business* (Blakiston, 1947), p. 413.

Price collection agencies are aware of this defect of their data, but there is no easy solution to the difficulty.[28] Industrialists have a major stake in correcting this situation, for data which showed changes in actual transaction prices would be more useful to them, and they would appear better before the public. Their prices would, in most cases, be found to be more flexible. It seems possible, therefore, that industry could be persuaded to report realizations particularly if they were not tied to narrow product specifications.[29] Certainly industry shares with governmental agencies the responsibility for good price statistics. But if sellers will not report realizations, then it may be necessary to turn to buyers. In some commodity lines brokers may be an easier and more accurate source than are either sellers or buyers. Regardless of the data sources, I suggest that in the wholesale markets for many commodities getting transaction rather than quoted prices is more important than being certain that either series refers to an item of exacting specifications. Furthermore for many commodities a very small sample of realizations would be preferable to a wide coverage of quoted prices.

(2) Dealing with the problem of changes in the representativeness or importance of commodities, or of market segments for a commodity line, within the concept of a pure price measure forces compromise with the availability of data. Where such changes are significant in the short run and the necessary weight data are available, I am in general agreement with Professor Bruce Mudgett's argument for use of year-by-year link indexes.[30] Our interest in current price movements is generally greater than in the comparison of today's prices with those of ten years ago. Precise, up-to-date weights may not be at hand but that leads me to observe that somewhat less than ideal facts for up-to-date weights are preferable to precise weights ten years old. No feasible degree of promptness in revising the items or channels of sales quoted, or the weights, will portray the very short-term product and sales mix shifts, however. From the viewpoint of measuring "pure" price change that simply means that weights cannot be revised that rapidly.

(3) This brings us to the problem which is most difficult conceptual-

[28] The Bureau of Labor Statistics participated in such tests of their data as that by Saul Nelson in the National Resource Committee's report *The Structure of the American Economy* (1939), Appendix 1, and then carried out for OPA the study on steel prices referred to in footnote 22.

[29] Industry may be concerned that such information for specific companies might leak to business rivals or to agencies enforcing the Robinson-Patman Act. Variations in realization relative to quoted prices suggest the presence of discriminatory pricing. This fear of industry would be lessened if they were asked to report realizations on a broader segment of their business than a carefully specified item.

[30] A summary of Mudgett's paper appears in *Econometrica*, Vol. 18, No. 3, July, 1950, pp. 289–90.

ly and perhaps in its data requirements. That is whether, or under what circumstances, the absolute level of price of items of changed importance should be permitted, or made to, affect the composite measure. The issue can be made clear by a simple example. Suppose two brands of a commodity sell at substantially different prices and that between two dates the transaction price and quality of neither brand changes. But assume, as happens not infrequently, there is a sizeable shift of volume away from the low-priced to the high-priced brand, and to simplify our case, that this occurs without a change in total number of units of both combined. Clearly aggregate dollars received by the sellers will increase and so will those paid out by the buyers. But no change in the total number of units, or of the price of either brand, will have occurred. Is the increase of aggregate value of product traceable to an increase of output or of prices? No increase has occurred in the number of pounds, yet a conventional production index using price per pound weights would show a rise in output. On the other hand, the average price paid per pound, disregarding brands, actually has increased. While this is not a "pure" price increase, it affects what sellers get per pound and what buyers must spend per pound which is important for some analytical problems. This is particularly true in all problems where it is important that the price index times the output index provide a measure of total value and that the relative importance of price and output in contributing to value changes be gauged accurately. Such are objectives 3, 4, and 5 outlined above on pp. 465–466.

What is needed in such cases is a device which divides changes in aggregate value into a "production"[31] factor and a "price"[31] factor. The former can best be measured by inputs, and this amounts to saying that for each period the weights of that period should be used.[32] Indeed the precedent for such a practice has already been set in the Federal Reserve Board's Index of Industrial Production where the output of about half of the industries (by weight) is measured by man-hours, and, in addition, in other industries by materials consumed.[33] In this

[31] These words are put in quotation marks because they are used out of their usual connotations.

[32] The issue posed here seems to be similar to that raised by Professor Mudgett in that he wants to use all of the price and quantity data of both periods. But he seeks to do this by a "cross-weight" index such as Fisher's "ideal" formula. That device merely reduces, but does not eliminate the measurement problem for which the device proposed here was designed. Furthermore, for markets of the type stressed here the weight data for Fisher's formula could not be obtained.

[33] For the purposes here these inputs would have to be more finely tailored to the industry than is true of data used in the F.R.B. production index. For example, for rubber tires, for which the F.R.B. measures output by man-hours, it would be better to use both man-hours and rubber consumed.

way the output measure reflects sales and product mix shifts or product redesign or dropping or adding items only to the extent that they alter inputs. The remainder of their effect on aggregate value, measured by dividing an aggregate value index by an input index, reflects effects of the varied conditions of price formation in different market segments. While this is not "price" in the ordinary meaning, it is what price-making forces bring about. Where the study is of these forces at work this equivalent of price is conceptually correct.

The procedure just described is carried out in a simple form in computing the "average hourly earnings index" and in "census" indexes of unit values.[34] In these, average unit values are obtained by dividing payrolls or value of output by number of hours worked or number of units produced and the quotients put in index number form. This method seems not to raise objections even though these indexes do not measure "pure price." Perhaps the acceptance stems from the availability of a physical common denominator for output such as for many food items.[35] But this would not help for machinery, furniture, or apparel for the very reasons which led the Federal Reserve to measure output by manpower input in such industries.

Therefore, a more general but conceptually identical method is needed in which output is measured by physical inputs. An example of how this is done, and of the usability of the measure follows: In late 1937 the tire industry advanced quoted prices three per cent in spite of continued depressed conditions following the sharp recession earlier in the year. One might suspect that this was a "paper" increase to be offset by concessions and the other competitive methods for which the industry at all levels is famous in times when volume is low.[36] But when the device proposed here was applied, "average realization per input unit" was found also to have advanced.[37] While the exact percentage found here cannot be defended because of some crudities in the data, it is significant that the quotient moved in the same direction as did the list prices.[38]

---

[34] Index compiled from census data are used to measure unit value changes in Solomon Fabricant's *Output of Manufacturing Industries, 1899–1939*, National Bureau of Economic Research (1940), and in A. C. Neal, *op. cit.*

[35] Thus the average value per pound of margarine produced, computed from census data, fell 6.8 cents from 1929–1931, but the reported wholesale price declined only 4.2 cents.

[36] Replacement market sales were as high in 1938 as in 1937, but original equipment sales were over 50 per cent lower in 1938.

[37] The details were as follow:

$$\frac{\text{net dollar sales in 1938 as a percentage of 1937}}{\text{man-hour and rubber inputs in 1938 as a percentage of 1937}} = \frac{=100.5}{=94.8} = 106.$$

[38] Furthermore, profits on replacement business were as high in 1938 as in 1937. Finally, one of the major producers examined their records for the author, and reported that the general level of realizations on tires were higher in 1938 than in 1937.

This procedure of using input to measure output and then of dividing such a measure of "production" into an aggregate value index amounts to assigning all of the movements of aggregate value not accounted for by output change to the "price" component.[39] Only to the extent that variations in physical inputs account for movements of aggregate value of an industry, or the economy, is it an output change. Changes in value because of money cost movements, widening or shrinking of profit margins, changes in what is produced, shifts in weights, changes in representativeness of items in a commodity field, to the extent that they do not call for greater or lesser amounts of physical inputs, are considered the "price" component of the change of aggregate value.[40] This is what is intended, for those industries and those analytical purposes stressed here. Predominantly this residual is the result of those attributes of markets and of their modes of adjustment which should be the focus of "price" research.[41]

To emphasize doubly the tentative character and limited usefulness of the device just described, four cautions are mentioned specifically. It does not measure prices in the ordinary meaning of that term and should be identified by some such term as "average value per input unit." Second, it is distinctly a limited-purpose measure, and should not be used where the objective calls for a measure of "pure price" change. Third, it is useful for short period comparisons only, chiefly because adjustments cannot be made for the effects of technology on inputs. But for the purposes for which such a measure is designed, short term comparisons are primary. Fourth, it falls down on the promptness criterion, and is useful only for historical research. Indeed there is great difficulty getting value and input data which match each other in coverage.

Even if the procedure proposed here is judged conceptually correct for certain price measurement purposes, it is so difficult to apply that often the same objectives will have to be reached by a study of what is shown by a variety of measures, which use different data and methods

---

[39] For purposes of constructing the measure proposed here the origin of the developments affecting average value per input unit is not important. But in explaining why the average value changes, or in pointing to the significance of such a movement, it is important to know whether developments, other than changes in transaction prices for a fixed bill of goods, take place on the initiative of buyers or of sellers. Even there, it is presumed that both parties acquiesce, and each is checked by the other except in periods when artificial controls exist under OPA ceilings.

[40] Although the method just described was designed for the mix-shift problems, it would also solve the transaction-price, weight-shift, and product-redesign problems. Net sales or value of output figures correctly reported should reflect price concessions and differential price performance of products of changed importance. Inputs should reflect effect on "output" of weight-shifts and the importance of product redesign.

[41] Actually such research deals with inter-relationships among product, volume, production costs, selling costs, and prices.

of construction but no one of which is itself a satisfactory gauge. Certainly one would not suggest that common weight indexes be dropped, for they are conceptually correct *for certain purposes* regardless of the structure of industries and of their methods of competitive adjustment. And for some other markets, such as those for raw materials, common weight measures are valid for most purposes. What is suggested is that no measure be considered all-purpose without examination along the lines indicated here and that progress be made in developing measure for those situations for which available price series and indexes are inadequate.

VI. SOME COMMENTS ON PUBLISHED INDEXES

This discussion of purposes and problems of measurement of price change leads to some observations about particular indexes now available. The Consumer Price Index is a useful guide in income distribution problems for the income groups to which it applies. Hence it may be an appropriate measure of equitable adjustments of basic wage rates as is provided under some current contracts and public policies. But to assume that workers' money income, even from straight time hours of work, will change by the same percentage as adjustment of basic rates under this formula, is to overlook what happens under either the public or private "administration" of wages. When demand for labor changes sharply, developments occur which are analogous to those sketched above with respect to commodity markets. The shift in occupational weightings, performance standards for jobs and other administrative decisions within a given wage structure can alter radically the average straight time hourly rate as the War Labor Board experience showed to an accentuated degree.[42] Consequently the income effects of wage changes, best indicated by average straight time hourly earnings, will ordinarily be greater than adjustments in basic rates made to conform with movements of the C.P.I. To the degree that this occurs, workers' real income before taxes per hour will advance, although those who develop the policy appear to think it will not.

Then the individual or group sub-indexes of the C.P.I. do not permit fine comparisons of the movement, say, of the retail prices of apparel or house furnishings with prices or factor costs at earlier stages

[42] From January 1941 to the same month in 1945, basic wage rates advanced 20 per cent, straight-time hourly earnings even after adjustment for inter-industry shifts by 37 per cent and average hourly earnings by 57 per cent. A significant amount of the spread among these wage measures occurred before wage stabilisation was imposed. (Figures summarized in *Business Week*, April 7, 1945, p. 19.)

of these industries. Not only is product coverage progressively less complete as one moves upward from the raw material stages, but also all of the modes of adaptation outlined above become more important. Such vertical comparisons are more accurate in the food field than in most others, but even there they lack the precision necessary in many cases for the "marketing margin" comparisons made.[43] Because only the "best-selling" item is priced by the reporters this provides neither an absolute measure of retail prices for all of the items or brands in the field that the item is presumed to represent nor does such an item price bear a constant relationship to those other item prices when a price movement or structural change occurs.

The C.P.I. is not a reliable measure of whether retail prices are out of line with consumers' incomes, although it has often been so used.[44] This is not primarily because of price concessions, in as much as retail price reporters usually get transaction prices except on "big ticket" items. To a greater degree the limited commodity coverage inhibits such use of the index, as the Bureau emphasizes. The major problem is that the index does not record, except on occasion, changes in what people buy—but what they buy is what is being produced and must be sold to maintain employment. But because the index has a major role in its present form it should not be revised to fit the uses discussed here. Instead some experimenting could be done with a census type[45] average value per physical unit of a category of goods to see how this compares with price changes for the representative item. An even broader measure of this sort is a chain store company's index of prices per ton of goods handled—computed from dollars sales divided by tonnage—which might be found to be indicative of what is happening in a form that would be useful for some purposes.[46]

Much of what was said in earlier sections was obviously pointed at wholesale price indexes. The only point to be added here has to do with the use of available data for specific industries. In no one of the several industries which I have examined intensively have I found these series satisfactory for the "analysis of the price structure of the economy and changing interrelationships among individual commodities," a use suggested by the B.L.S.[47] Even for purposes which call for less precision

[43] See R. O. Been, *Agricultural Information Bulletin No. 4* (Bureau of Agricultural Economics, 1949) for a compilation of such margins.

[44] As for example in the "Nathan Report" of December 11, 1946, which was summarized in the New York *Times* of the following day.

[45] See p. 474 above for description of such indexes.

[46] This index showed a 7 per cent rise from the third quarter of 1949 to the same period of 1950 whereas the food component of the C.P.I. advanced only 3 per cent.

[47] See the Bureau's *Bulletin* No. 993, p. 25.

one doubts whether such a claim should be made for such semi-manufactured industrial materials as castings or upholstery fabrics, and even more so for the products thereof, such as machinery, mechanical rubber goods, or furniture. These are the areas where the complex of price concessions, weight changes and price differences among substitutes are so characteristic that published price data can give a quite false view of the facts whether one is interested in changes in pure price or in average value per input unit.

One finds the components of the farm parity index to be in part inept for the purposes for which that index is designed, but good for others. The whole concept is a welfare one; hence prices paid should be for a fixed bill of goods, and particularly for goods used in production, should measure usefulness per dollar spent. Yet the data reported for inclusion in the prices-paid index may reflect changes in the quality or price line of a commodity when there occurs such a change in what is "commonly" bought by farmers. Thus, the mere advance of farm incomes, if it affects qualities bought, may raise the parity index even though no price changes for industrial items have occurred. To this degree the index is not a welfare measure,[48] but is structurally more indicative of farmers' ability to buy at existing prices those goods which they choose to buy than is the C.P.I. with reference to urban workers' incomes. Perhaps even more important is the assumption that the truck or tractor of 1930, 1940 and 1950 are the same in utility.[49] It is true that the same assumption is made about the prices received for wheat and apples, but technological change in raw materials has been far less than for manufactured goods, particularly for those made of metals. Consequently the parity index has a built-in relative gain for the farmers. Granted that the use of such indexes is a policy, not a statistical, decision, nevertheless it is one made possible by, and is perhaps open to influence by, those who construct indexes.

VII. SOME CONCLUDING COMMENTS

As an appraisal this has been largely critical, in part of those who gather, manipulate and report price changes, but more so of the "grab them and use them" practice of consumers of price statistics. One applauds the professional competence and immunity to pressure of public and private workers in this field. But their task is never done

[48] The extent to which shifts in what is bought affects the index is not known. A study was made under abnormal war time conditions, but this does not provide a basis for generalization. A more typical comparison would be between qualities, say of men's shirts, bought in 1939 and in 1949, for in the latter year supplies were again liberal and quality choices were those of the consumers.

[49] This is not to suggest that there is an easy way to measure "usefulness." But it is to say that policy conclusions, based on such indexes, are to an extent open to question.

for the phenomena they deal with change in character and so do the purposes of those who use price facts.

Our resources for price work will always be limited; consequently both redirection where necessary and conservation of resources devoted to price work are called for. There should be a more exhaustive comparison than has been possible here of current index number practice with what we need to know about changes of a price. A similar investigation is needed of the problem of measuring changes of a price sort in specific industries. All duplicating or overlapping price collecting and index number construction activities should be re-examined to see whether really important additions of fact, promptness, or accuracy are obtained. This would not preclude private checking by occasional studies of the performance of public price-collecting agencies, or vice versa. But except for that, effort which essentially duplicates that done by another group had better instead be directed to developing needed special-purpose measures, among which should be those which would portray more accurately developments in a commodity area or at a functional level.

These comments on the division of effort in price work, and the earlier critique of existing indexes, lead to two suggestions about procedure. The first is that the agencies, public or private, engaged in price work should encourage, indeed sponsor, research on price measurement problems by persons outside their respective groups. One phase of such research, that of measuring price developments in specific industries of complex structure, should be a promising field for doctoral candidates, provided that industry cooperation is obtained. Second, agencies releasing index numbers should both prepare and publicize statements of what their indexes do and do not do, and of the types of inferences which can correctly be drawn thereform. A notable start in this direction has been made by the B.L.S.[50] Such statements are needed to warn the unwary, which I suspect includes a large part of the economics profession.

In both research on price measurement, and in using what is at hand, I stress that price measurement calls for adaptation of methods to the varied markets recorded and to uses to which the measures are put. To imply that one type of measure can meet all of these needs and situations is to assume a "plain and simple doctrine as to the relation between cost of production, demand, and value," which Marshall warned us "is necessarily false."[51]

[50] See *Bulletin* 993 referred to above. The discussion of uses and limitations of the respective indexes are quite brief, and may not fully advise the non-specialist. In addition the present author would claim less for the wholesale index and particularly for sub-indexes and individual series than does the Bureau.

[51] Alfred Marshall, *Principles of Economics* (1920), p. 368.

MATHEMATICAL MODELS IN PSYCHOLOGICAL SCALING*

C. H. COOMBS**
University of Michigan

THE major theme of this paper is to stress the importance to social science and psychology of the application of mathematical and statistical theory applicable to numbers which are not elements of fields.[1] The greater part of statistical theory is concerned with numbers to which the operations of arithmetic can be validly applied. There is reason to believe that in social and psychological data some the axioms of arithmetic are not satisfied and, if this is the case, weaker axiomatic systems, like partially ordered sets (posets), should be used to provide the logical framework for the description and analysis of tbe data. The nature of some of these weaker systems, their experimental basis, and the reasons for their importance will be discussed.

I. A DISTINCTION BETWEEN MEASUREMENT AND SCALING

In current psychological literature there are several different meanings given to the term psychological scaling. Because the term will be used in a very particular sense in this paper, a different meaning sometimes given to the term will be discussed in order to make clear the sense in which I shall use it. To some people, psychological scaling means the measurement of psychological magnitudes which have no apparent physical correlate. In this sense the measurement of qualitative attributes, such as attitudes and mental abilities, constitute problems of psychological scaling, e.g., the degree to which an individual is liberal or conservative, the measurement of leadership ability, the degree to which some statement of opinion is pro-x, the amount of arithmetic ability an individual possesses, or how popular some child is in a group of children.

The term psychological scaling is used in this sense in distinction to psychophysical scaling where the psychological attribute has a known physical correlate, as in the case of the perceived heaviness of weights, the brightness of lights, and the length of lines. This distinction arises

* The research on which this paper is based was made possible by a grant from the board of Governors of the Horace H. Rackham School of Graduate Studies, University of Michigan.

** I wish to express my appreciation and indebtedness to many helpful criticisms and suggestions of my colleagues Professor R. M. Thrall, Dr. Howard Raiffa, and Mr. Donald Mela of the Department of Mathematics. They, of course, have no responsibility for the accuracy or adequacy of the contents of this paper.

A modified version of this paper was presented before the session on Statistical Problems and Psychological Theory, at the meeting of the American Statistical Association, December 28, 1950.

[1] All of the operations of arithmetic are permissible on numbers which are elements of fields.

because there have been two serious rational attempts to solve the fundamental problem of psychological measurement, one requiring the measurement of a physical correlate and the other being independent of the measurement of a physical correlate.

While the two solutions to this problem need not be presented in any detail, some of their general characteristics will be pointed out. The fundamental issue which both of these rational approaches face is the problem of a unit of measurement. Fechner's solution to the problem is known as Fechner's Law [9]. He accepted Weber's relation as a law and went a step further. Weber, somewhat over a hundred years ago, observed that the increment which had to be added to a stimulus to be just noticeable (say 75% of the time) seems to bear a constant ratio to the magnitude of the stimulus. Fechner defined these just-noticeable differences to be equal psychological units, and from this he derived a logarithmic relation between the psychological magnitude and the physical magnitude of a stimulus.

Psychological measurement by means of Fechner's Law requires the measurement of a physical correlate in the stimulus,[2] the presentation of stimuli which are relatively indiscriminable, and the assumption that equally-often-noticed differences are equal. One criticism that is made of Fechner's rationale is that any functional dependence of a psychological unit of measurement on the measurement of a physical attribute is at the outset assuming a solution to one of the problems we seek to study.

Thurstone, in his Law of *Comparative* Judgment [8], provides an alternative rational solution to the basic problem of psychological measurement. Obviously, most psychological magnitudes which we desire to measure have no measurable physical correlate, and, even if we admitted the assumptions underlying Fechner's Law, this rationale is not widely applicable. Thurstone provides a solution to the problem which is not dependent upon a measurable physical correlate. He assumes that the psychological response magnitudes aroused by a given stimulus are normally distributed on a psychological continuum, and he substitutes this assumption for the measurement of a physical attribute. Thurstone's approach is to use the standard deviation of this normal distribution as the unit of psychological measurement. This approach, however, still requires judgments with relatively indiscriminable stimuli and its Case V still makes the assumption that

[2] The technique of scale construction by fractionation appears also to be dependent upon an underlying continuous physical correlate. One does not have a continuous simple order of statements of opinion about X such that one can make the judgment "this statement lies midway between these two statements."

equally-often-noticed differences are equal. Thurstone's solution constitutes a generalization of Fechner's psychophysical solution, both as a method of collecting data and as a method of analyzing data. To characterize psychological scaling, then, by contrasting it with psychophysical scaling is merely to recognize a historical distinction and not a logical distinction.

There is, however, another basis on which psychological scaling may be characterized, and that is the logical basis of a scale. In general, there are two aspects to every type of scale. There is, on the one hand, an abstract or formal system consisting of elements, operations, and properties. On the other hand, there is a system of real objects with operations that may be performed on them and with properties experimentally observable. Measurement, in a broad sense, is the process of mapping the object system into an abstract system. It behooves us, then, to map the object system into an abstract system the axioms of which are satisfied by the object system, otherwise there is no meaning in the object system for the operations performed in the abstract system. We shall use the axiomatic basis of the various scales to differentiate between those we shall classify as psychological measurement and those we shall classify as psychological scaling. This is not a conventional distinction but one which is desirable for my immediate purposes.

Most people, when first asked to define "measurement," will answer with the statement that measurement is the assignment of numbers to objects. When pushed a little further, it becomes apparent that there are still considerable differences in what they mean. Some people are satisfied when the numbers merely identify classes of objects—this constitutes a mapping between the numbers and equivalence classes of objects. The result of such a use of numbers is called a nominal scale, and, of course, any other set of distinguishable symbols will serve as well as numbers.[3] This is the level of measurement which is approached by psychiatric classification systems.

Some people would specify that the term "measurement" implies the relation "greater than" between the numbers. If this condition, among others, is satisfied in the object system, then this kind of measurement is called an ordinal scale. This type of scale is typified by a scale of hardness based on the operation of "A scratches B."

If we further require that the operations of addition and multiplication on the numbers be permitted in the sense that these operations in

[3] Cf. also (6), (7).

the abstract system correspond to operations on the object system and hence have meaning, then we have what is called a ratio scale. This type of scale is typified by the measurement of length and weight.

There is another type of scale falling between an ordinal and a ratio scale which has been put to considerable use in psychology and the social sciences. This is the interval scale. In the case of the interval scale the operations of arithmetic are permissible on the *differences* between numbers assigned to objects rather than on the numbers themselves. This is the type of scale presumed to hold in the Centigrade or Fahrenheit scales of temperature.

To summarize the discussion up to this point it may be said that the interval and ratio scales have in common a constant unit of measurement, but differ in that the ratio scale has an absolute zero and the interval scale an arbitrary zero. The ordinal scale has no unit of measurement and no origin but has the quantitative property of transitivity of the relation "greater than."

Because of the prevailing attitude that many of the operations of arithmetic are permissible on numbers which are measures of something, I shall use the term measurement to apply to the use of ratio or interval scales and the term scaling to the other kinds of scales such as ordinal and nominal scales.[4] Generally speaking, I am using the term "scaling" to mean measurement without the use of numbers, where measurement values must be expressed in partially ordered sets or posets rather than in the real numbers.

It is immediately evident why the psychologist and social scientist have striven to achieve at least interval scales. Once he has a unit of measurement that is constant within and between individuals, he has measurement by an interval scale, and there are available to him strong mathematical and statistical systems for the description and analysis of data and the construction of theory. For example, all statistical theory is applicable to measurement by a ratio scale, and very little is lost if one has only an interval scale rather than a ratio scale. In particular the coefficient of variation is the only index that comes to my mind as applicable to a ratio scale and not to an interval scale. This is not a serious loss.

Because of the availability of strong mathematical and statistical systems, the social scientist is tempted to assume that his data satisfy the necessary postulates, and to convert his data into numbers which

[4] It should be clear that this distinction constitutes an arbitrary line made for purposes of convenience of discussion.

are elements of an interval scale. He then must reject such portions of the data that are incompatible with this preferred abstract system. What we sometimes fail to recognize is that the measurement theory assumed in analyzing data becomes a part of those data, and such portions of the data which are incompatible with the a priori preferred abstract system are rejected and regarded as constituting error variance.

II. THE ORDERED METRIC SCALE

A distinction has been made between psychological scaling and psychological measurement on the basis of the axiom systems that underlie them. In particular, psychological measurement refers to measurement by means of interval and ratio scales, and psychological scaling refers to scales built on axiom systems which do not possess a constant and universal unit of measurement. Two types of scales which fall into the class of psychological scaling so defined are the ordinal and the nominal scale, both of which have been discussed. There are, in addition, at least two more non-numerical kinds of scales which can be recognized experimentally, namely, the ordered metric and the partially ordered scale. The axiom system of the ordered metric falls between the ordinal and the interval scale, and the axiom system of the partially ordered scale falls between the ordinal and nominal scale.

Any abstract type of scale, to be anything more than a branch of mathematics, must have an image in the object system, and there must be operations in the object system which give meaning to the abstract system. The experimental operations underlying the ordered metric have been presented in an earlier paper [3]. In brief, this type of scale is one in which the objects are simply ordered, and hence constitute an ordinal scale, but in addition the distances between objects are at least partially ordered. This type of scale falls logically between an ordinal scale and an interval scale but is in the class of psychological scaling rather than psychological measurement as distinguished here.

An experimental operation or technique for arriving at an ordered metric has been called the unfolding technique and has two characteristics which distinguish it from the conventional procedures of psychological measurement. In the first place, the stimuli which are used need not be relatively indiscriminable but may be 100 per cent discriminable.[5] This is an advantage. The second distinguishing characteristic

[5] One of the referee's of this paper suggests that this be called macroscaling to contrast with microscaling in which relatively indiscriminable stimuli are used. He points out that microscaling may reveal a simply ordered scale which "may not be unidimensional in the large due to twists and turns in a scale which is imbedded in an essentially larger space" (private communication).

will be regarded by some as an advantage, by others as a disadvantage. This distinguishing characteristic is that the data one obtains do not always constitute a set of I scales which unfold into a single J scale; in fact, they rarely do. In other words, the technique does not always work. If one *sets out* to construct an ordered metric, one may be bitterly disappointed. This method permits behavior to exhibit its complexity, and a unidimensional solution will not be obtained if the behavior is multidimensional. I think of this technique as a device for *finding* scales rather than *making* scales. The validity of constructing a scale to measure attitude toward x and carrying it around to measure other groups of people is a question of experimental fact. The attribute attitude toward the negro, for example, may be *qualitatively* different for different cultural groups. The psychological implications of this have been discussed elsewhere [2].

What has been said up to this point about psychological scaling has had reference, at least implicitly, to unidimensional scaling, to the case where stimuli and/or individuals may be positioned on a single continuum. But the generalization of these ideas to multidimensional unfolding seems possible. In the unidimensional case the I scale can be thought of as picking up a continuum like a piece of string at the locus of the individual and as giving the rank order of the stimuli in terms of their relative distance away from the individual in either direction. In the multidimensional case, it is not a length of string but a multidimensional surface which is being picked up at the locus of the individual, and the I scale gives the rank order of the stimuli in terms of their relative distance away from the individual in either direction. In the two-dimensional case it would be like the individual picking up a handkerchief at his locus on that surface and collapsing it into his I scale, giving the stimuli in rank order, from the nearest to the farthest, regardless of direction. This generalization is being developed by Mr. Joseph Bennett, whose Ph.D. thesis [1] provides several different bases for determining the dimensionality of a set of rank orders. These developments are the beginning of the analogue of Thurstone's multiple-factor analysis, but for qualitative data with no unit of measurement assumed.

To summarize briefly up to this point, a distinction between psychological measurement and psychological scaling has been made on the basis that psychological measurement involves the use of scales which permit the operations of arithmetic on the numbers assigned to objects, in particular the ratio and interval scales. Psychological scaling, on the other hand, is constituted of those scales not possessing a

constant and common unit of measurement, including the ordered metric, the ordinal, the partially ordered, and the nominal scales. All these have been described except the partially ordered scale, which will now be approached circuitously.

III. THE PARTIALLY ORDERED SCALE

When one speaks of attempting to construct a scale, one usually means locating stimuli and/or individuals on a unidimensional continuum. It is sometimes known that such continua are really continua in a multidimensional space and that the attribute in question is really a function of two or more primitive or underlying latent attributes. The conventional mathematical model which is applied to construct the unidimensional continuum in a multidimensional space I call "the summative model." In this model the components underlying a given type of behavior are assumed to compensate for each other, to combine additively in the construction of the observed behavior. This model is exemplified by the multiple regression equation. For example, if I may be permitted to oversimplify for purposes of illustration, an individual's grade in a course may be regarded as a linear combination of the individual's intelligence and application to his studies. Two individuals of varying intelligence may get the same grade because the less intelligent one studies more.

But there may be some behaviors for which the primitive attributes do not compensate but act in a manner I shall call conjunctive behavior. By this I mean behavior in which several attributes are required, *each* to a minimum degree. Again, if I may oversimplify for the sake of illustration, supervisory ability may be constituted of an ability to deal with people and a knowledge of the job, but each of these is required and they do not compensate for each other. Thus, some particular job element to be performed successfully may require that the individual possess a certain minimum amount of knowledge about the job *and at the same time* a certain minimum amount of ability to deal with people. In this model, lack of either primitive attribute means failure, and surplus of one does not compensate for lack of the other.

This type of behavior, to my knowledge, was first described by H. M. Johnson in 1935 [4] but has since been neglected. This type of behavior has an obvious counterpart in the mathematical model of what can be called disjunctive behavior, in which the primitive attributes are combined on an "either-or" basis. In this instance the model says that a particular task may be performed successfully by an individual with

the minimum amount of one attribute *or* a minimum amount of some other attribute. Thus, if we were trying to scale individuals on an attribute that we call "success in life," such an attribute might be disjunctive in that different individuals may achieve "success in life" by virtue of entirely different underlying attributes. If one attempts to scale individuals in behavior which is either conjunctive or disjunctive, the relation "greater than" will hold for some pairs but not necessarily all pairs. The result is a partially ordered scale. The problem then is to decompose the partially ordered scale into sets of simple orders which are psychologically meaningful. This will also be recognized as being an analogue of multiple factor analysis. These developments in psychological scaling, while exploratory and incomplete, have been partially successful through the application of lattice theory and topology. These are the branches of mathematics which social scientists have most seriously neglected and the ones which appear particularly promising.

These new developments of psychological scaling suggest certain problems of particular interest to mathematical statisticians. The development of non-parametric statistics is an obvious "must" for the description and analysis of data in psychological scaling. Then also, there are new problems which appear. In conventional correlation theory, for example, we always correlate two continua, though each may be a summative composite in a multiple dimensional space. But in this new approach there is the suggestion that some attributes cannot be represented on a line but must be represented by surfaces of two or more dimensions, as for example, the attributes of supervisory ability and success in life. Another example which I will again oversimplify might be musical ability. This ability may be thought of as a two-dimensional surface with the primitive attributes of a sensitive ear and finger dexterity. Then musical ability of an individual would be the conjunction of these primitive abilities, and the one does not compensate for the lack of the other. Such attributes would be abstract spaces and may be without even a reference frame. Yet it would be of value to have an analogue for correlation between attributes which can only be represented by such abstract spaces. For example, it would be desirable to have an index to measure the concept of correlation between two posets.

IV. THE DILEMMA OF THE SOCIAL SCIENTIST

Up to the present I have introduced no concepts of error, probability, or sampling in the theoretical structure. Certainly, at some point, and perhaps the sooner the better, probability concepts, and certainly error

concepts, must be introduced. My avoidance simply reflects a dislike for having the freedom to classify parts of the data as error at will. This provides the experimenter or theoretician with a blank check, and he is able to pay almost any price to find evidence in his data for certain preconceptions with which he approaches the data. In my opinion, it is too easy to arrive at sufficient conclusions when there is no way of estimating whether the conclusions are necessary.

Consider any attribute, for example attitude toward the British. We may assume there is such an attribute qualitatively the same with the same meaning for each member of a group of individuals who differ only in their degree of being pro or anti British or we may ask whether there is such a unidimensional attribute in the first place.

If we test whether the attribute exists, as even an ordinal scale, by either Guttman's scaling technique or the unfolding technique we may find it does not. But, and this is an important point, we may still assume that it does, and by other measurement techniques, construct an interval scale and "measure" the degree to which each individual is pro or anti British.

This again illustrates the difference between approaching the data with a weak and a strong measurement theory. The difference between the two approaches is a difference in the degree to which the integrity of the data is maintained. A weak system leans more heavily on the data, a strong system will yield a solution in spite of some of the data. One way of looking at the difference between these approaches is the answer they give to the question "what part of the data will be called error?" This problem is the dilemma of the social scientist.

The social scientist, preferably before collecting his data, but usually after, has as his first problem the selection of a scale of measurement for expressing the data. Usually this has meant converting the data into numbers on an interval scale so that the powerful system of arithmetic becomes available for the reduction, description, and analysis of the data. In selecting a scale of measurement he is in effect making the axioms of that type of scale his assumptions or postulates which are then imposed on the data. If he chooses a strong axiom system, like a ratio or interval scale, he will put more into the data, he will have powerful mathematical and statistical tools available, he will be more likely to get a solution and it will be a simpler one. On the other hand he will have more error variance, he will fit less of his data, and what he gets out will to a greater degree represent what he has put into the data by his assumptions.

If he chooses one of the weaker systems he will have less error vari-

ance, he will fit more of his data, and he will be able to test his postulates more readily. On the other hand, he will not have elaborate statistical models for testing hypotheses (yet) and he will get more complex answers. In fact, he will find that behavior is rarely very simple.

I don't think that there is a solution of this dilemma that will satisfy everyone. Each of us, as a result of our individual value systems and the indoctrination we have had, will choose accordingly. As a matter of fact I would be inclined to state that the choice is in part dependent upon the objectives of the research, and for some purposes one should use a strong system and for other purposes a weaker one, although I frankly admit to a strong bias for a weak system.

I would like to close this paper with a quotation from Frank Knight [5], an economist, who describes the prevailing measurement procedures in the social sciences with the dictum "If you can't measure, measure anyhow." The spirit of this dictum, to assume away your difficulties, is the exact antithesis of modern scaling theory.

REFERENCES

[1] J. F. Bennett, *The Dimensionality of a Set of Rank Orders*, Ph.D. Thesis, University of Michigan, 1951.

[2] C. H. Coombs, "Some Hypotheses for the Analysis of Qualitative Variables," *Psychological Review*, 55, 167–74, 1948.

[3] C. H. Coombs, "Psychological Scaling Without a Unit of Measurement," *Psychological Review*, 57, 145–58, 1950.

[4] H. M. Johnson, "Some Neglected Principles in Aptitude Testing," *American Journal of Psychology*, 47, 159–65, 1935.

[5] F. H. Knight, " 'What is Truth?' in Economics," *Journal of Political Economy*, 48, 1–32, 1940.

[6] S. S. Stevens, "On the Theory of Scales of Measurement," *Science*, 103, 677–80, 1946.

[7] S. S. Stevens, Editor, *Handbook of Experimental Psychology*, Chapter 1. Mathematics, Measurement, and Psychophysics, S. S. Stevens. New York: John Wiley & Sons, Inc., 1951.

[8] L. L. Thurstone, "A Law of Comparative Judgment," *Psychological Review*, 34, 273–86, 1927.

[9] R. S. Woodworth, *Experimental Psychology*. New York: Henry Holt and Company, 1938.

YATES' CORRECTION AND THE STATISTICIANS

FRANZ ADLER*
University of Arkansas

I

THE social scientist as a rule lives under the shadow of a cloud. He pities himself because his field is not an "exact" science. And he tries to gain exactness by introducing figures into his discourse, preferably figures with many decimals, arrived at by the use of the most complex and time consuming statistical procedures available. While quantification is a laudable effort and, if correctly performed, does add to the exactness of our knowledge, it should be a comfort for every social scientist to know that figures and exactness are not the same thing. Even the field of statistics at times shows all the disconcerting features which have given the social sciences their bad reputation.

This fact is to be illustrated here by the case of Yates' correction for continuity as applied to 2×2 tables.

Yates' correction is intended to diminish the discrepancy between results achieved by the use of the exact distribution[1] and by the chi-square method. The latter method is faster, simpler, and therefore preferable. The chi-square distribution, however, is continuous, while the distribution it is supposed to approximate is discontinuous. This discrepancy can be diminished: "If we group the chi-square distribution, taking the half units of deviation from expectation as the group boundaries, we may expect a much closer approximation to the true distribution. This is equivalent to computing the values of chi-square for deviations half a unit less than the true deviations, 8 successes for example being reckoned as $7\frac{1}{2}$, 2 as $2\frac{1}{2}$."[2]

Notice that Yates does not indicate specifically how a research worker applying his correction is to determine which ones of the four figures in his square he is to raise and which ones he is to lower by $\frac{1}{2}$. A mathematician's reply to such a comment is, of course, that anybody reading Yates' paper with any understanding of what is said there, will know. The trouble with this argument is that it is empirically false.

No less a man than R. A. Fisher, whose name is never mentioned without awe and reverence in circles of initiates, fails to give his readers

* Special thanks for help and advice are due to Mrs. Leta M. Adler, Instructor in Statistics at the Institute of Science and Technology of the University of Arkansas.

[1] See R. A. Fisher, *Statistical Methods for Research Workers*, eleventh edition, Hafner Publishing Company, New York, 1950, Sec. 21.02.

[2] F. Yates, "Contingency Tables Involving Small Numbers and the χ^2 Test," *Supplement to the Journal of the Royal Statistical Society*, Vol. 1 (1934), p. 222.

a clear account of Yates' correction. Instead of speaking of the reduction of *deviations* by a half a unit, he recommends "entering the table of χ^2, not with the value calculated from the actual frequencies, but with the value it would have if our *observed frequencies* had been less extreme than they really are each by half a unit."[3] Probably misled by this presentation, and not enlightened by his own reading of Yates' original article, Rider, a respected statistician and textbook author, advises his readers to apply the correction by "adding ½ to the smallest frequency of the table and adjusting the others so that the marginal totals will remain the same."[4]

In most cases, Rider's error will be of no consequence. The smallest figure in a four-fold table will very often be one of the two which are below the expected value, whose deviation is negative and is consequently decreased by adding ½. But it can (and does) happen that the smallest number is larger than its expected cell frequency while the second smallest number is smaller than its expected cell frequency. In this case, adding ½ to the smallest number would increase the error instead of correcting it. Table I shows an example.

TABLE I

	Observed values				Expected values		
	I	II			I	II	
a	16	74	90	a	18	72	90
b	4	6	10	b	2	8	10
	20	80	100		20	80	100

The observed value of 4 is the smallest value in the table. It is, however, larger than the expected value of 2 which corresponds to it. The deviation of the observed value from the expected value would be increased instead of decreased if we followed Rider in adding ½ to it. The value which must be increased by ½ in this table is 6. It corresponds to an expected value of 8, and by adding ½ to the observed value, the deviation is decreased. The deviation in the first case can be decreased only by subtracting ½ from 4. Thus, Rider's instructions, based on Fisher's wording which refers to extreme frequencies instead of to deviations, and which in turn is based on Yates' lack of clear instructions will, in cases of this kind, cause the trusting research worker to do exactly the opposite of what he should do.

[3] R. A. Fisher, *l.c.*, p. 94; in his example on p. 95, Fisher indicates once more that he does not think of deviations, but of actual frequencies: " . . . we should rewrite the table with the larger frequencies 10 and 15 reduced by a half, and the smaller frequencies 2 and 3 increased by half." The italics are the present author's.

[4] Paul R. Rider, *An Introduction to Modern Statistical Methods*, John Wiley and Sons, Inc., New York, 1939, p. 113.

This point was later discussed by Cochran who expressed the rules to be followed in general terms similar to those used by Yates.[5] He does not indicate which value in the square should be reduced, which one increased, but leaves this matter again to the dubious ingenuity of the reader. The best statement published on how to proceed is the one by Paulson and Wallis: "First, adjust the table of observations so that each frequency is brought one half unit nearer the frequency expected if there were no long run difference between the two methods [under study]. This may be done by adjusting one frequency and then changing the others so that the marginal totals of the new table equal those of the old."[6] Thus, 13 years after Yates' original publication, a practical, clear instruction on how to proceed on this point finally became available.

In his example of the application of Yates' correction to a fourfold table, Fisher drops a hint. He points to the fact that *in this particular example* the corrected difference of the cross product of the table is half the total number of observations less than the uncorrected value. He does not claim that this is always the case, much less does he offer the proof that subtraction and addition of $\frac{1}{2}$ in the prescribed manner will always lead to this result.[7] Quinn McNemar follows this hint and gives his readers only one formula for the corrected chi-square:

$$(1) \qquad \chi^2 = \frac{(|AD - BC| - N/2)^2 N}{(A + B)(C + D)(A + C)(B + D)}$$

where A, B, C, D are cell frequencies and N is the total frequency.[8] He does not offer any proof of the generality of Fisher's observation either. Neither has the present writer found proof anywhere in the literature he consulted. Such proof is, however, possible.[9]

But is it true that this subtraction of $N/2$ *always* produces the desired result, i.e. a corrected chi-square, a chi-square more closely approximating the true value than the uncorrected one? Let us assume that the difference of the cross products be zero. In this case, the correction which is supposed to yield a lower chi-square value in all

[5] W. G. Cochran, "The Correction for Continuity," *Iowa State College Journal of Science*, Vol. XVI, #4, July 1942, p. 423.

[6] Edward Paulson and W. Allen Wallis, "Planning and Analysing Experiments for Comparing Two Percentages," Chapter 7 in Statistical Research Group, Columbia University, *Techniques of Statistical Analysis*, McGraw-Hill Book Company, Inc., New York, 1947, p. 254. The authors stress the point that this method applies to the use of χ^2 in fourfold tables only. Bracketed insert ours.

[7] R. A. Fisher, *l.c.*, p. 95.

[8] Quinn McNemar, *Psychological Statistics*, John Wiley and Sons, Inc., New York, 1949, p. 200.

[9] The writer relies here upon his colleague, Professor Raymond H. Burros, of the Psychology Department of the University of Arkansas, who has developed such a proof.

cases[10] actually yields a higher value; any real number squared is more than zero. Or let us assume that the difference of the cross products is smaller than one half of $N/2$, that is than $N/4$. Again the correction will bring an increase rather than a decrease in the numerical value of chi-square. This may lead to values which may seem to indicate significance, if not on the one or five per cent level perhaps so at the twenty or thirty per cent level.[11]

The question may also be raised here, whether aside from these cases, formula (1) is adequate for all applications of the chi-square technique. It is important to distinguish between cases in which one is testing against a one-sided or a two-sided alternative, and if it is a one-sided one which side it is. As it stands, the formula covers the two-sided alternative, that is the hypothesis that a difference as large as that of the cross products, whether it be positive or negative ($|AD-BC|$), is due to chance. In the case of the one-sided alternative, however, only one of the following hypotheses is tested: either that a difference as high (algebraically) as this one is due to chance, or that a difference as low (algebraically) as this one is due to chance.

In cases where the distribution of chi is supposed to approximate an exact distribution of a fairly symmetrical shape, it will be all right to disregard the direction of a one-sided alternative and to accept the probability from one tail of the normal curve. A possible discrepancy between χ and the exact distribution will be about the same on either tail. Where the exact distribution is asymmetrical, however, this procedure may lead to over-estimation at one tail of the curve, to under-estimation at the other. It is then that the direction of the one-sided alternative becomes important. While in the case of the two-sided alternative, over- and under-estimation at the two ends may balance more or less, the one-sided alternative does not allow for any balancing.

Yates, while undoubtedly aware of the difference between one-sided and two-sided alternatives, does not give any specific rules on how to proceed differently in the two cases. He merely points to the fact that his correction does not improve the situation with regard to "violently unsymmetrical" distributions, but he still believes that the "ordinary test corrected for continuity" will avoid consistent over- or under-estimates of significance.[12] Perhaps "ordinary" in this context means "with a two-sided alternative"; perhaps it doesn't. Again, a patient

[10] Cf. for example George W. Snedecor, *Statistical Methods, Applied to Experiments in Agriculture and Biology*, The Iowa State College Press, Ames, Iowa, 1940, p. 170.

[11] Chi-squares up to the 50% level are used by George Peter Murdock in his *Social Structure*, The Macmillan Company, New York 1949, p. 129 *et passim*.

[12] F. Yates, *l.c.*, pp. 223 and 229.

mathematician might reason out what Yates did mean or should have meant. The fact is that no text the present writer has encountered contains a clear statement of how to recognize fourfold tables which belong to very asymmetrical distributions and are therefore not improved by the use of Yates' correction. No text contains a clear statement on how to proceed differently with regard to one-sided and two-sided alternatives. No text warns the reader on how to use the above formula in cases of one-sided alternatives.[13]

Lack of clarity and explicitness regarding limitations to the usefulness of the correction are the rule rather than the exception. Thus, neither Yates' original article nor any other published source known to the author does take into account the possibility that addition and subtraction of $\frac{1}{2}$ might increase rather than reduce the value of chi-square. Table II shows a constructed example of such a set of data:

TABLE II

	Observed values				Expected values		
	I	II			I	II	
a	1	2	3	a	0.9	2.1	3
b	2	5	7	b	2.1	4.9	7
	3	7	10		3	7	10

In this example the deviation of the observed value of 1 from the expected value of 0.9 is 0.1. Proceeding according to prescription, we subtract $\frac{1}{2}$ from 1; thus we achieve a "corrected" observed value of 0.5, and instead of reducing the deviation, we have increased it from 0.1 to 0.4. Since the same is true for all four values in the square, the chi-square is increased by this application of the correction rather than decreased. This will occur whenever the difference between observed and expected frequencies is smaller than 0.25.

It is obvious that the distortion produced by proceeding in this manner or by using formula 1 when $|A-B| < N/4$—these two methods lead, of course to the same results—is worst when the data are smallest and gets successively more and more negligible with an increase of the frequencies. But it is also obvious that the correction has most effect with small frequencies.

[13] James G. Smith and Acheson J. Duncan, *Sampling Statistics and Applications, Fundamentals of the Theory of Statistics*, McGraw-Hill Book Company, Inc., New York 1945, pp. 395–397, discuss one example of a one-sided alternative, without ever generalising as to rules of procedure or interpretation in such cases. Cochran, *l.c.*, p. 423, mentions one-sided alternatives in passing.

II

So far this paper has dealt with Yates' correction in its primary function as a continuity correction, that is, its function of making a continuous distribution better approximate what is really a discontinuous distribution. The rest of the paper will deal with the problem of small expectations and the use of the correction in overcoming it.

These are some of the question involved: Are there any minimum observed or theoretical cell frequencies beyond which the chi-square test becomes inapplicable? Is there a minimum N (sum of all the cell frequencies) beyond which the test becomes inapplicable? Are these limits or others those beyond which Yates' correction must be applied to achieve meaningful results? While most writers on the subject answer the first question, they do not necessarily answer the second or third. Nevertheless, these questions seem almost inseparable and, for this reason, are treated here together.

While all of the inaccuracies mentioned previously may be considered errors of individuals rather than of the profession, as absence of a discovery which still had to be made (by the present writer, it must be presumed), or, more probably, as matters so obvious to the expert that there never seemed to be any need to tell about them to those mathematical laymen who were supposed to use these techniques, there appears in the matter of limiting values such a basic lack of agreement and such prevalence of divergent subjective judgments as have become rare today in those fields of social inquiry to which scientific methods are being applied.

Yates, who as the originator of the correction under discussion might be assumed to know what he is after, suggests the use of a special table whenever expectancies are below 100, but believes that "on the other hand the worker will not be led badly astray if he applies the ordinary chi-square test (after correcting for continuity) to tables giving expectations as low as 10, so long as the corresponding distributions are reasonably symmetrical. Moreover, in view of the fact that the discrepancies are of opposite signs on the two tails of the distribution, no consistent under- or over-estimate of significance will be made when applying the ordinary test corrected for continuity to a heterogeneous collection of data, even with very small expectations."[14] Whether

[14] F. Yates, *l.c.*, p. 229; as stated previously, Yates does not refer to one-sided alternatives here. But even in cases of two-sided alternatives, it is only small comfort that the discrepancies do not produce "consistent" over- or under-estimates. The individual estimate, if false, remains false, even if the next case would balance it. Corrected chi-squares cannot validly be added, so that even the use of numerous fourfold tables does not correct for the error of each single computation. It would, indeed, be preferable to have a consistent over- or under-estimation so as to be able to expect an error in a given direction rather than to have to wonder whether attained results are too high or too low.

"very small expectations" are above or below 10 is not stated in general; an expectation of 1.74 in an example is said, however, to be "very small."[15] In the same context, Yates offers an example in which two of the expected frequencies lie between 2 and 3. He solves this problem by three methods and states definitely that corrected chi-square gives an excellent approximation to the result achieved by the exact method.[16] He does not demand a minimum total frequency (N), but requests that his correction be used whenever the smallest expectation is less that 500.[17] He gives as his reason the fact that omission of the correction may at the 2.5% point of either tail of the distribution amount to as much as 0.26%.[18]

Fisher's approach is confusing. He points to the fact that the case of the fourfold table is different from other cases of chi-square application in that there is only one degree of freedom. In other cases he stipulates that in no group shall the expected number be less than 5. In this special case, he says, we can deal with small observed numbers like 3 or 2, provided we use Yates' correction.[19] There seems to be, however, no contrast between *expected* numbers above 5 and *observed* numbers below 5. The two may well occur in the same cell of a fourfold tables a they do in Fisher's own example.[20] Since Fisher does not give a generalized rule of operation, it may well be that some readers might get the impression that he suggests Yates' correction is to be employed if an expected cell frequency is below 5 while others may decide that it is to be employed if an observed cell frequency is below 5.

It is probably on the strength of such readings that Rider advises his readers to use Yates' correction whenever "the number in one or more compartments of the table is small, say less than 5"[21] and that McNemar tells his readers that it "should always be used when any one E is less than 5 and is advisable when an E is less than 10."[22] McNemar's E refers to an expected frequency, while Rider seems to be talking about observed frequencies.

Rider believes—though unjustifiably so—that there is support for his views in Irwin's statement: "When . . . numbers in all cells are small, the nonapproximative method should be used, but if samples are of reasonable size (say giving 30 individuals or more when clubbed

[15] *Ibid.*, p. 232.
[16] *Ibid.*, p. 231.
[17] *Ibid.*, p. 230.
[18] *Ibid.*, p. 229.
[19] R. A. Fisher, *l.c.*, pp. 93 f.
[20] *Ibid.*, p. 95.
[21] Paul R. Rider, *l.c.*, p. 112.
[22] Quinn McNemar, *l.c.*, p. 207.

together) and there are small frequencies (i.e. yielding expected frequencies <5) in one or two cells owing to the percentage of marked individuals being 90 per cent or more, we shall seldom be misled by applying the usual test and taking the 1 per cent level of significance . . . "[23] Rider may have overlooked the parenthetical reference to expected frequencies and just noticed the term "small frequencies" as such. He may also never have seen the complete article and may have read an abstract or excerpt of it instead; he quotes it as occupying pp. 1–94,[24] while actually it occupies only pp. 83–94 of the volume. More important still, when using the term "usual test" Irwin may be referring to the uncorrected chi-square procedure, though this cannot be stated with certainty. In a table preceding the quoted lines, he shows under the heading "Usual Method" uncorrected values and then, in brackets, corrected values. In a footnote appended to the title of the article, he states: "This paper was concluded in May 1933, but its publication has been unavoidably delayed. Meanwhile a paper dealing with the same subject, in some respects more completely, has been published by F. Yates ("J. Roy. Stat. Soc." Suppl. I, 2, p. 217). Yates' correction to the usual method of calculating χ^2 consists in deducting one half from the deviations between observed and expected frequences. The effects of applying it is [sic] shown in Table X of this paper, it will be seen as a rule to give good agreement with method (3) at the critical levels."[25] Whether or not any changes other than the footnote and the bracketed values were introduced into the paper to adapt it to the changed scientific situation cannot be learned from it. Whether then the quoted lines refer to the unchanged, uncorrected values gained by the "usual method" or to the bracketed values gained by applying the correction, or to both sets of values perhaps, must remain a matter of conjecture.

A rather large number of statisticians seem to feel that Yates' correction is not a means of overcoming the objections raised against the use of chi-square with very small cell frequencies. Edwards is willing to accept theoretical frequencies not lower than 5 provided the correction is being used.[26] Cramér demands 10 as a minimum theoretical cell frequency.[27] Hoel names theoretical frequencies "somewhat greater

[23] J. O. Irwin, "Tests of Significance for Differences Between Percentages Based on Small Numbers," *Metron*, vol. XII, no. 2, May 1935, p. 94.

[24] Paul R. Rider, *l.c.*, p. 112, footnote.

[25] J. O. Irwin, *l.c.*, p. 83, footnote.

[26] Allen L. Edwards, "On 'The Use and Misuse of the Chi-Square Test'—The Case of the 2×2 Contingency Table," *Psychological Bulletin*, Vol. 47, No. 4, July 1950, p. 342, *et passim*.

[27] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, 1946, p. 420 f.

than 5" as his minimum for the fourfold table.[28] Guilford accepts 5 as a minimum value but considers 10 as better. But he suggests that "when any of the theoretical cell frequencies are small (some say as small as 50, in the opinion of the author 25 might be a more realistic limit), it is recommended that . . . Yates' correction . . . be applied."[29] According to Snedecor "no expected number should be smaller than 5, and when possible all should be 10 or more . . . It is well worth making [the correction for continuity] if any critical decision is involved, especially if the numbers in any class fall below 50."[30] Yule and Kendall demand that N, the sum of all cell frequencies, be "reasonably large . . . It is difficult to say exactly what constitutes largeness, but as an arbitrary figure we may say that N should be at least 50, . . . No theoretical cell frequency should be small. Here again it is hard to say what constitutes smallness, but 5 should be regarded as the very minimum, and 10 is better."[31] Lewis and Burke tend to hold out for 10 as the minimum figure for theoretical cell frequencies to which a chi-square test can usefully be applied.[32]

Finally there is Peters who has the following to say: "There is much talk about what minimum number of frequencies is required in a cell before using it in a chi-square calculation: whether five or ten or twenty or what. The number required for *exactness* in fulfilling the assumptions is infinite; for *any* number short of infinity the determination is only rough, but good enough. . . . "[33]

It would seem strange that all this controversy should go on with no attempt to settle it. Most of the participants try to do this by either citing authorities at each other or by illustrating their arguments with bits of mathematical reasoning which show that their points hold for certain cases, two methods which seem to leave the mutual opponents quite unimpressed. But one would expect that some proof would be attempted that could settle the argument for everybody. Edwards attempted an empirical solution. He computed a set of four examples of corrected chi-squares with theoretical cell frequencies

[28] P. G. Hoel, *Introduction to Mathematical Statistics*, John Wiley and Sons, Inc., New York, 1947, p. 191.

[29] J. P. Guilford, *Fundamental Statistics in Psychology and Education*, second Edition, McGraw-Hill Book Company, Inc., New York, 1950. pp. 278, 279.

[30] George W. Snedecor, *l.c.*, pp. 168, 170. Bracketed insert ours.

[31] G. Udny Yule and M. G. Kendall, *An Introduction to the Theory of Statistics*, Charles Griffin & Company, Limited, London, 1937, p. 422.

[32] Don Lewis and C. J. Burke, "The Use and Misuse of the Chi-Square Test," *Psychological Bulletin*, Vol. 46, No. 6, November 1949, pp. 462, 463, 486.

[33] Charles C. Peters, "The Misuse of Chi-Square—A Reply to Lewis and Burke," *Psychological Bulletin*, Vol. 47, No. 4, July 1950, p. 336 fn.

below 5 and he came to the conclusion that the probabilities achieved were quite compatible with those obtained by a direct method.[34] But neither he nor Lewis and Burke seem impressed by this evidence: they decide to "wait and see" until a mathematical proof is forthcoming.[35] They seem to forget that in the meantime they owe the mathematical proof that their restrictions are valid.

Federighi has recently published a paper reporting construction of tables for all possible fourfold chi-squares with total frequencies up to 40. This choice of 40 is due to his belief that an expected cell frequency of ten is necessary to achieve acceptable results by the usual method. His tables are based on almost 5000 applications of Fisher's formula for the exact treatment of fourfold tables[36] in a simple but ingenious modification. All the user of the tables has to do is to find the following four figures: the smallest row or column sum, the second smallest row or column sum, the observed cell frequency this row and this column have in common, and the total frequency. Entering the table with these four figures, he can immediately determine whether he is to accept or reject his hypothesis on the five percent, one percent, or one tenth of one percent level. Unfortunately only a small sample of these tables was published in the original article.[37] The tables seem to be easy to handle and not too voluminous. They certainly merit inclusion in future statistical texts or handbooks.

It may well be that the solution to the problem of chi-square with small frequencies lies not in further speculations concerning the limits of the applicability of Yates' correction, but in the expansion of Federighi's computations to the point that anybody having reasonable doubts about the smallness of his sample and the applicability of the usual corrected or uncorrected formula can use them. So far, only chi-square with $N \leqq 40$ can be looked up in these tables.

The tables are not made to find exact probabilities. If these statistics are wanted, all the previously discussed difficulties still persist.[38]

---

[34] Allen L. Edwards, *l.c.*, p. 345. *Cf.* also Table X of Irwin's article mentioned above.

[35] Don Lewis and C. J. Burke, "Further Discussion of the Use and Misuse of the Chi-Square Test," *Psychological Bulletin*, Vol. 47, No. 4, July 1950, p. 355.

[36] R. A. Fisher, *l.c.*, pp. 96 f.

[37] Enrico Federighi, "The Use of Chi-Square in Small Samples, Methodological Note on Over-Estimation of Relationship in 2 ×2 Chi-Square Tables Where N is Fewer Than or Equal to 40 and One or More of the Expected Cell Frequencies is Fewer than 10," *American Sociological Review*, Vol. 15, #6, December 1950, pp. 777–779.

[38] The research worker finding himself in this difficulty, may wish to refer to Paul G. Hoel, "On the Chi-Square Distribution for Small Samples," *The Annals of Mathematical Statistics*, Vol. 9 (1938), pp. 158–165. He may then decide to use the usual method after all or to abandon chi-square for some other less disputed method.

III

It is doubtful whether any housewife would buy a cookbook knowingly which contains in the majority recipes which are well reasoned but which have not been tested in practice.[39] She buys the book if she feels reasonably certain that the operations indicated therein actually do yield the sauce or pudding they claim to yield. The proof of the pudding is in the eating, not in the logical computation that garlic and chocolate, each tasting fine separately, ought to taste still better together. The buyer of a statistical text also should be assured, whenever possible, that the recipes he finds there are not merely logically sound, but also tested by practical application to facts of experience.

The mathematical statistician can—and often does—provide logical soundness.[40] It is the business of the research worker who applies the statistical methods to his field to report which one of alternative methods he used brought results most closely consistent with facts of his experience. It is he who should report whether his generalizations or predictions based on corrected chi-squares were more satisfactory than those based on uncorrected chi-squares, whether his labors led to satisfactory results when he used chi-square with low frequencies or low expectations or low totals or whether they didn't. Such reports could be forthcoming in the social sciences only if verification were practiced more instead of merely being preached. Accumulating experience would then enable the textbook writer to supplement the results of mathematical logic with the results of continued experimentation. The realities of any field are not of necessity identical or parallel to mathematical assumptions. Only in a mutual exchange of ideas and experiences between the mathematical statistician and the practical re-

[39] Two shelves of cook books in the Agricultural library of the University of Arkansas were inspected with regard to this remark. Of 38 books all but 4 contained an easily noticeable reference to testing. They stated on the cover, on the title page, in the preface, in the foreword, or in the dedication that the recipes were tested by the author or her informant, that the author was a cook of many (preferably fifty) years of experience (or her mother, aunt, grandmother, or cook was) and that the book contained the recipes which her family loved, that the recipes were the secrets of the great and famous chefs of the past or present, or they contained other similar remarks intended to convince the buyer of the book that the recipes contained therein do work.

[40] Contrary to popular belief, mathematicians, particularly statisticians, are not completely aloof with regard to empirical (as contrasted to logical or mathematical) experimentation. Compare for example G. Udny Yule and M. G. Kendall, *l.c.*, p. 439. The number of experiments needed for valid assertions is, in a case like the present one, unfortunately so tremendous as to be out of the question. It is perhaps for this reason that Sukhatme's experiments (P. V. Sukhatme, "On the Distribution of χ^2 in Samples of the Poisson Series." Supplement to the *Journal of the Royal Statistical Society*, Vol. V, 1938, pp. 75 to 79) have remained without decisive influence upon the above mentioned discussion. By studying 1100 random samples of numbers he came to the conclusion that even if the expected value in a cell is as low as one, the test of goodness of fit "is tolerably good when the sample size is 15." He does not state whether or not his findings are valid for the fourfold table as well as for higher degrees of freedom.

search worker can it be found which assumptions, which methods, and which procedures fit which field.

Another demand a housewife may make when buying a cookbook is that it be written in a language she understands. But the preceding discussion, particularly in the first part of this paper, has shown a tendency of statisticians to express themselves so cryptically that even their own colleagues don't understand them. A buyer of a text in statistics has a right to be given clear, unambiguous, and complete instructions on what operations to use under what conditions. Inaccuracy in the use of the common tongue may undo all the exactness gained by the manipulation of mathematical symbols and figures. Not every research worker has the time and ability to return to original sources, to compare the opinions of the authorities, or to derive his own mathematical formulas and procedures. If he is able and desirous to do all this he does not need a text. But if he needs one, he is entitled to expect from it at least as much as a housewife expects from her cookbook, granting merely that cookery as a field is somewhat older than statistics, and that the latter deserves some leniency on that account.

THE DISTRIBUTION OF THE QUOTIENT OF RANGES IN SAMPLES FROM A RECTANGULAR POPULATION

PAUL R. RIDER
Washington University and Flight Research Laboratory (U.S.A.F.)

THIS paper gives the distribution of the quotient of the ranges of two independent, random samples from a continuous rectangular population and discusses the features and properties of the distribution curve for various sample sizes. The distribution is independent of the population range and can be used to test the hypothesis that two samples came from the same rectangular population, just as the distribution of the variance ratio is used to test whether two samples came from the same normal population.[1]

Introduction. The population which we wish to consider is the rectangular population given by

$$f(x) = \begin{cases} 1/R_0 & \text{for } 0 = x = R_0, \\ 0 & \text{elsewhere.} \end{cases} \tag{1}$$

If R is the range of a random sample of size n from this population, then the distribution of R is

$$n(n-1)R_0^{-n}R^{n-2}(R_0 - R)dR. \tag{2}$$

This is a well known result; see [1], page 210.

To derive the distribution of the quotient of the ranges of two independent, random samples from the population (1) we let R_1 be the range of a sample of m and R_2 the range of a sample of n. The joint distribution of R_1 and R_2 is

$$m(m-1)n(n-1)R_0^{-m-n}R_1^{m-2}R_2^{n-2}(R_0 - R_1)(R_0 - R_2)dR_1dR_2, \tag{3}$$

R_1 and R_2 varying between 0 and R_0. From this we wish to determine the distribution of $u = R_1/R_2$.

Distribution of the quotient of ranges. It is convenient to regard R_2 as the abscissa and R_1 as the ordinate in rectangular coordinates, and to transform to polar coordinates by setting $R_1 = r \sin\theta$, $R_2 = r\cos\theta$. Then (3) becomes a function of r and θ, which can readily be reduced to a

[1] The quotient of two average ranges in samples from a normal population has been used in industrial quality control in estimating the variance ratio (see, for example, [2], pages 70–72). In an article [3] which appeared since this paper was submitted for publication, the distribution of the ratio in samples from normal populations has been discussed.

function of r and u, since $u=\tan\theta$. The variable r can be integrated out in the usual manner, giving the distribution of u, which is as follows:[2]

$$\frac{m(m-1)n(n-1)}{(m+n)(m+n-1)(m+n-2)}\left[(m+n)u^{m-2}-(m+n-2)u^{m-1}\right]du,$$

$$0\leqq u\leqq 1, \quad (4)$$

$$\frac{m(m-1)n(n-1)}{(m+n)(m+n-1)(m+n-2)}\left[(m+n)u^{-n}-(m+n-2)u^{-n-1}\right]du,$$

$$1\leqq u<\infty. \quad (5)$$

It will be noted that the distribution is entirely independent of the range R_0 of the population.

It will also be noted that (4) and (5) can be obtained, the one from the other, by replacing u by $1/u$, interchanging m and n and prefixing a minus sign. This must of course be the case, since the distribution of $1/u$ is that of R_2/R_1.

As (4) and (5) are readily integrated, it is easy to find the probability that the ratio R_1/R_2 will lie between any specified values u_1 and u_2 or that it will exceed a specified value. (This is much simpler than finding the value that will be exceeded a specified proportion of times in random samples.) Thus the hypothesis that two samples came from the same rectangular population can be tested.

Description of the distribution curve. The distribution curve obviously has the positive u-axis as an asymptote. Although it is composed of two parts, these parts are joined together smoothly at the point whose abscissa is 1. In fact, at the junction both parts have not only equal ordinates but also equal first and second derivatives. They always differ in their third derivatives, however.

If $m=2$ the left-hand part of the curve is a straight line which starts from a point $2(n-1)/(n+1)$ units directly above the origin and has the slope $-2n(n-1)/(n+2)(n+1)$. This slope has the value $-1/3$ when $n=2$ and approaches -2 as n increases.

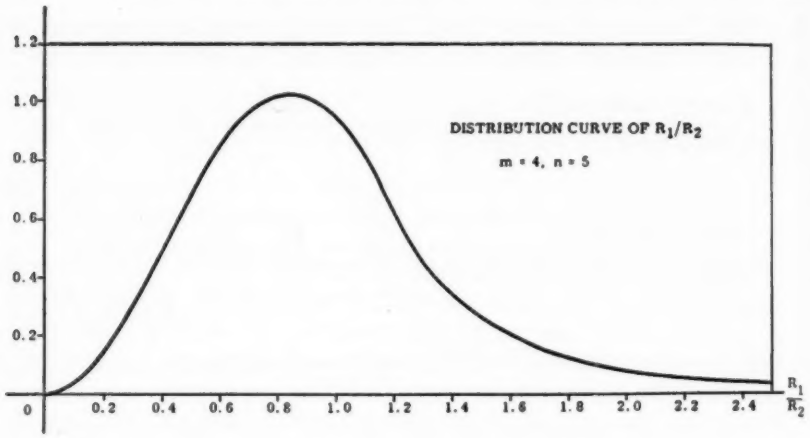
If $m=3$ the curve starts from the origin with the slope $6n(n-1)/(n+2)(n+1)$. This slope has the value 1 when $n=2$ and approaches 6 as n increases.

If $m>3$ the curve has a zero slope at the origin.

A graph of the curve for the case $m=4$, $n=5$ is shown in the accompanying figure.

---

[2] I am indebted to the Editors for pointing out that the distribution given in [3] is incorrect.



Mode and inflection points. The mode of the curve is given by

$$u = (m - 2)(m + n)/(m - 1)(m + n - 2) \tag{6}$$

or by

$$u = (n + 1)(m + n - 2)/n(m + n) \tag{7}$$

according as $m-n$ is less than or greater than 2. If $m-n=2$, the mode is given by the common value of (6) and (7), namely, $u=1$.

The left-hand inflection point is given by

$$u = (m - 3)(m + n)/(m - 1)(m + n - 2), \quad (m > 3) \tag{8}$$

the right-hand inflection point by

$$u = (n + 2)(m + n - 2)/n(m + n), \quad (m > 2). \tag{9}$$

Median. The median will be less than, equal to, or greater than 1 according as m is greater than, equal to, or less than n.

Moments. The kth moment about the origin is

$$\mu_k' = m(m-1)n(n-1)/(m+k)(m-1+k)(n-k)(n-1-k). \tag{10}$$

In particular, the mean is

$$\mu_1' = (m - 1)n/(m + 1)(n - 2). \tag{11}$$

It follows that the mean will be less than, equal to, or greater than 1 according as $n-m$ is greater than, equal to, or less than 1.

Obviously all moments of order greater than $n-2$ are infinite.

If m and n approach ∞, μ'_k approaches 1. Consequently μ_k, the kth moment about the mean, approaches 0. This is intuitively evident, for as m and n increase, the values of R_1 and R_2 will converge in probability to R_0, the population range, and the quotient R_1/R_2 will approach 1. That is, the distribution of u will approach $f(u)=1$, $u=1$, whose mean is 1 and whose moments about this mean are all 0.

Samples of equal size. Many of the foregoing formulas simplify somewhat if $m=n$. For example, (10) becomes

$$\mu_k' = n^2(n-1)^2/(n^2-k^2)\left[(n-1)^2-k^2\right]. \tag{12}$$

In the preceding section it was stated that μ_k' approaches 1 and consequently μ_k approaches 0 as the sizes of both samples approach infinity. Thus the values of $\alpha_3=\mu_3/\mu_2^{3/2}$ and $\alpha_4=\mu_4/\mu_2^2$ become indeterminate forms. Nevertheless, for the case $m=n$ at least, $\alpha_3\to 0$ and $\alpha_4\to 4\frac{1}{2}$ as $n\to\infty$. For, employing the usual formulas for changing moments about the origin to moments about the mean, we find, after a considerable amount of simplification,

$$\mu_2=\frac{4n^2(n-1)^2}{(n+1)^2(n-2)^2}\cdot\frac{1}{(n+2)(n-3)},$$

$$\mu_3=\frac{4n^2(n-1)^2}{(n+1)^3(n-2)^3}\cdot\frac{21n^2-21n-2}{(n+3)(n+2)(n-3)(n-4)},$$

$$\mu_4=\frac{4n^2(n-1)^2}{(n+1)^4(n-2)^4}\cdot\frac{6(3n^6-9n^5+103n^4-191n^3+76n^2+18n-4)}{(n+4)(n+3)(n+2)(n-3)(n-4)(n-5)},$$

$$\alpha_3=\frac{(n+2)^{1/2}(n-3)^{1/2}(21n^2-21n-2)}{2n(n-1)(n+3)(n-4)},$$

$$\alpha_4=\frac{3(n+2)(n-3)(3n^6-9n^5+103n^4-191n^3+76n^2+18n-4)}{2n^2(n-1)^2(n+4)(n+3)(n-4)(n-5)}.$$

It follows at once that $\alpha_3\to 0$, $\alpha_4\to 4\frac{1}{2}$ as $n\to\infty$.

Some idea of the rapidity with which α_3 and α_4 approach their limiting values may be gained from table 1.

TABLE 1

$n=m$	α_3	α_4	$n=m$	α_3	α_4
6	2.85	36.68	10	1.23	8.40
7	2.10	17.96	20	0.55+	5.23
8	1.68	12.39	100	0.11	4.53

Tables of percentage points. Brief tables of values of the quotient of ranges which will be exceeded in 10%, 5%, and 1% of random samples are given as tables 2, 3, 4, respectively.[3]

Example. The following data, selected from [4], page 173, are measurements of the width of a slot in a certain airplane part. The unit is 0.001 inch over 0.800 inch.

Sample 1 (Series 1): 77, 80, 78, 72, 78.

Sample 2 (Series 7 and 8): 75, 77, 75, 76, 77, 79, 75, 78, 77, 76.

Sample 1 was taken on the first day of production, sample 2 was taken two days later after the process had had an opportunity of becoming stabilized.

The hypothesis which we lay down is that both samples came from the same rectangular population. (A rectangular population might conceivably be a reasonable assumption if specification limits were narrow and only measurements between specification limits were retained.) We have

$$R_1 = 80 - 72 = 8, \quad m = 5; \quad R_2 = 79 - 75 = 4, \quad n = 10; \quad u = R_1/R_2 = 8/4 = 2.$$

Reference to table 4 shows that this is above the 1% point, and we should consequently reject the hypothesis.

If it is desired to find the exact probability of obtaining a quotient as large as or larger than 2, we substitute $m = 5, n = 10$ in (5) and integrate from 2 to ∞. The result is 0.0013.

TABLE 2

10 PER CENT POINTS OF THE DISTRIBUTION OF THE QUOTIENT OF RANGES FROM A RECTANGULAR POPULATION

n	m = number in sample having greater range								
	2	3	4	5	6	7	8	9	10
2	6.41	9.69	11.7	13.0	13.9	14.6	15.1	15.6	15.9
3	2.00	2.75	2.89	3.50	3.72	3.89	4.02	4.12	4.21
4	1.36	1.77	2.02	2.19	2.31	2.40	2.48	2.54	2.59
5	1.13	1.43	1.60	1.72	1.81	1.87	1.93	1.97	2.01
6	1.02	1.25	1.39	1.48	1.55	1.61	1.65	1.68	1.71
7	—	1.15	1.27	1.35	1.41	1.45	1.49	1.52	1.54
8	—	1.09	1.19	1.26	1.31	1.35	1.38	1.41	1.43
9	—	1.05	1.14	1.20	1.24	1.28	1.30	1.33	1.35
10	—	1.01	1.10	1.15	1.19	1.22	1.25	1.27	1.29

[3] The author wishes to express his sincere thanks to Lieutenants Brian T. Hastings and Howard D. Allen, U.S.A.F., for computing these tables. I.B.M. equipment was used.

TABLE 3

5 PER CENT POINTS OF THE DISTRIBUTION OF THE QUOTIENT OF RANGES FROM A RECTANGULAR POPULATION

n	m = number in sample having greater range								
	2	3	4	5	6	7	8	9	10
2	13.1	19.7	23.4	26.3	28.2	29.6	30.7	31.6	32.3
3	2.94	4.00	4.64	5.08	5.39	5.63	5.82	5.97	6.09
4	1.79	2.31	2.62	2.83	2.99	3.11	3.20	3.28	3.34
5	1.41	1.75	1.96	2.10	2.20	2.28	2.34	2.40	2.44
6	1.23	1.49	1.64	1.75	1.83	1.89	1.94	1.98	2.01
7	1.12	1.34	1.46	1.55	1.61	1.66	1.70	1.73	1.76
8	1.06	1.24	1.35	1.47	1.47	1.52	1.55	1.58	1.60
9	1.02	1.17	1.27	1.33	1.38	1.42	1.45	1.47	1.49
10	—	1.13	1.21	1.27	1.31	1.34	1.37	1.39	1.41

TABLE 4

1 PER CENT POINTS OF THE DISTRIBUTION OF THE QUOTIENT OF RANGES FROM A RECTANGULAR POPULATION

n	m-number in sample having greater range								
	2	3	4	5	6	7	8	9	10
2	66.4	99.7	119.6	133.0	142.5	149.6	155.2	159.6	163.2
3	6.83	9.26	10.7	11.7	12.4	13.0	13.4	13.7	14.0
4	3.23	4.11	4.65	5.02	5.29	5.49	5.66	5.79	5.90
5	2.23	2.73	3.04	3.25	3.41	3.53	3.62	3.70	3.77
6	1.79	2.14	2.35	2.50	2.61	2.69	2.76	2.81	2.86
7	1.56	1.82	1.98	2.09	2.18	2.24	2.29	2.34	2.37
8	1.44	1.62	1.75	1.84	1.91	1.97	2.01	2.04	2.07
9	1.31	1.49	1.60	1.68	1.74	1.78	1.82	1.85	1.87
10	1.25	1.40	1.49	1.56	1.61	1.65	1.68	1.71	1.73

REFERENCES

[1] Neyman, J., and Pearson, E. S., "On the use and interpretation of certain test criteria for purposes of statistical inference. Part I," *Biometrika*, Vol. 20A (1928), pp. 175–240.

[2] Rodgers, J. W., "Statistical techniques associated with S.Q.C.," *Six Lectures on Statistical Quality Control*, Ministry of Production Birmingham District Production Committee, Quality Control Panel, Birmingham, England, 1944, pp. 67–74.

[3] Link, Richard F., The sampling distribution of the ratio of two ranges from independent samples," *Annals of Mathematical Statistics*, Vol. 21 (1950), pp. 112–6.

[4] Grant, E. L., *Statistical Quality Control*, McGraw-Hill (1946).

THE MEASUREMENT AND PREDICTION OF ATTITUDES*

WILLIAM G. MADOW
University of Illinois

THIS book contains the first comprehensive treatment of the Guttman and Lazarsfeld techniques in scaling. On the one hand it extends the important researches begun in this area by Thurstone and Likert some twenty years ago. On the other it breaks important new ground by giving formulations of the problem of attitude measurement that can be mathematically treated. This volume is not only a summary of past research, but also, and more important, it has led and will lead to valuable future work.

Many variables that are considered to be important in the social sciences are underlying or latent variables; they cannot themselves be measured but determine the values taken on by the variables that can be measured, the so-called manifest variables. Some illustrations are "the price level," "the business cycle," "business confidence," "economic level," "political sentiment," "intelligence," "attitude" and the like. Underlying variables occur also in other fields; for example "quality" of manufactured products.

The underlying variables in the social sciences are usually only vaguely or, at least, non-operationally defined. When, therefore, measurement with respect to these variables is attempted, the task is more difficult and complicated than measuring a manifest variable or an operationally defined underlying variable. If the underlying variable is operationally defined to be a vector each of whose components is a manifest variable, many difficulties disappear. This operational definition is often possible in studies of quality of manufactured goods, but in the social sciences the ability to define a vector of manifest variables to be the underlying variable is often questionable.

Scaling techniques are efforts to deal with the measurement of underlying variables, either by assigning measures of these variables or by ordering a population with respect to them. They are, therefore, of considerable importance to statisticians. Those interested in this area will find a great deal of important information in the volume being reviewed.

The book is essentially in two parts. Of these, the first, consisting of Chapters 2 through 11, is an excellent discussion of the Guttman and Lazarsfeld scaling techniques as of the time the book was written. (Considerable research has been completed since then.) The second, consisting of Chapters 13 through 16, contains two case studies in prediction. An illuminating discussion by Stouffer (chapters 1 and 12) introduces each part.

Guttman's and Lazarsfeld's scaling techniques

An extensive treatment of Guttman's approach to scaling is given in Chapters 2 through 9. Guttman (Chapter 2) discusses the general problem

* A review of *Measurement and Prediction* by Samuel A. Stouffer, Louis Guttman, Edward A. Suchman, Paul F. Lazarsfeld, Shirley A. Star, and John A. Clausen.

of attitude and opinion measurement. Here, among other things, is Guttman's point of view on the definition of an attitude, including his suggestion "to consider an attitude to be a delimited totality of behavior with respect to something." In the following chapter, Guttman discusses the "Basis for Scalogram Analysis." After first specifying an attitude or opinion universe by stating the "delimited totality of behavior" to which it pertains, Guttman accepts the dictum of Murphy, Murphy, and Newcomb that "no scale can really be called a scale unless one can tell from a given attitude that an individual will maintain every attitude falling to the right or left of that point" (Guttman says that he uses the word "item" where they use "attitude"). It follows from this characterization of a scale that knowing a person's rank in the scoring of the items of a scale will indicate exactly which responses he gave to each of the items, thus achieving a perfect coding of the items of the scale, i.e. the ranks are sufficient statistics. This is one of Guttman's equivalent definitions of a scale. Once the perfect scale is thus defined, the questions become: "How does one test whether the scale exists?" and if it does "How does one calculate the scale values?" Guttman has given answers to both of these questions in the coefficient of reproducibility and other tests of existence of a scale and the scalogram technique. In Chapters 4 and 5, E. A. Suchman presents a beautifully clear discussion of both the tests of the existence of a scale and the technique of scalogram analysis including the scalogram board and illustrations of scales and quasi-scales. Thus, Chapters 3, 4, and 5 constitute a unified discussion of the scalogram technique and cover many important matters in addition to those mentioned above. In Chapter 6, Guttman discusses the relation of scalogram analysis to other techniques of scaling, including the older techniques of Thurstone and Likert, as well as the Lazarsfeld technique presented later in this volume.

Chapter 7 (Suchman) considers the "Intensity Component in Attitude and Opinion Research." Guttman has suggested that this component be used to obtain a "cutting" point on attitude scales, since, presumably, the intensity with which one holds an attitude will be greatest at the extremes of the scale and lowest at the point where a small shift in attitude would change one from favorable to unfavorable. Not only is a non-mathematical discussion of this point of view included in Chapter 7, but also a mathematical discussion is included in Chapter 9, entitled "The Principal Components of Scale Analysis," and prepared by Guttman. The latter chapter contains both a non-mathematical and mathematical discussion of some of the implications of a perfect scale. In Chapter 8, Guttman discusses the problems of the reliability of scales, taking up the sampling of people, of items or questions, and of trials.

In Chapters 10 and 11, Lazarsfeld presents the "Logical and Mathematical Foundation of Latent Structure Analysis" and the "Interpretation and Computation of Some Latent Structures." As is the practice in this volume, a complete non-mathematical treatment is included in addition to the mathematical treatment. The essential point in Lazarsfeld's approach is that atti-

tude is assumed to be an underlying variable that determines the probability distribution of the possible responses to the items on the scale. Thus, the estimation of attitude is put into the standard form of statistical problems, and our whole apparatus of identification, estimation, and testing of hypotheses becomes available for use.

In these chapters, Lazarsfeld has presented essentially the first general mathematical model of a scale. He assumes the existence of one or more underlying variables. Furthermore, the values of the underlying variables determine the probability distribution of the responses to the items of which the scale consists. Then, he finds it convenient to introduce the notion of a "pure" test, i.e. one in which the probability distribution is such that the conditional responses to the items are independent for each person as well as among persons. He makes some simplifying assumptions concerning these conditional probability distributions that result in a solvable practical problem that is often sufficiently close to reality to be useful. He also discusses the relation of latent structure analysis to other techniques of scaling, and gives some detailed illustrations of the computation of scales by this technique.

Comments on the Guttman and Lazarsfeld approaches to scaling

1. *Problem formulation and mathematical models.* Both Guttman and Lazarsfeld have formulated the *problem* of scaling in a manner that enables them to construct mathematical models to which general statistical theory can be applied. This is a tremendous advance which will make it possible for theoretical statisticians to make contributions to this area. This reviewer is not very much impressed with statements that existing statistical theory does not deal with qualitative variables. What has been lacking, and, to a large extent, still is lacking in the social sciences, is *the formulation of realistic statistical problems involving qualitative variables.* For scaling, this lack no longer exists because of the Guttman and Lazarsfeld formulations, and we can confidently expect that important advances in scaling will result. Guttman and Lazarsfeld have, of course, gone much further than the formulation of problems. They have constructed the corresponding mathematical models and made considerable progress in solving the technical problems. Had they given only their formulations of the problems of scaling, we would have been indebted to them; our present debt is much the greater for their further work.

2. *Tests of the existence of a scale.* Guttman's definition of the perfect scale does not include the probability distribution of the responses to the items, except in the limiting cases, where certain responses must occur, and the others are impossible. The complete realization of this limiting case will but rarely occur. Consequently, Guttman's tests are essentially empirical rules of behavior that state when the deviations from the "null" hypothesis are small enough for us to act as though it is true in practice. It is not likely that these tests can be improved much within the present framework of Gutt-

man's model. Rather than criticize this lack, it is this reviewer's opinion that, for understanding the ideas of scaling the perfect scale is of the utmost value.

3. *Principal components of scales.* It is important to notice that one can have a perfect uni-dimensional scale and yet there may exist several principal components of the scale. This is not mathematically surprising. In mathematics, a curve in a plane can be given by its equation, or by specifying its x and y coordinates as functions of a third variable t, thus obtaining the parametric equations of the curve. The principal components of scales are parametric equations for the underlying variable. Hence, the only question is whether the parametric equations add or suggest anything. Guttman has made fruitful suggestions on the psychological identification of the first few principal components, and it is to be hoped that this work will continue. It is not essential, however, that "intensity" always be the second component. So far, this identification is based on the fact that the second principal component and empirical measures of intensity are U-shaped. Other psychological aspects of attitude may also be U-shaped. There is no question, however, but that the derivation of the principal components is yielding important results on the structure of attitudes.

4. *The definition of attitude and the sampling of items.* Let us denote the underlying variable that it is desired to measure by attitude$_1$. (Our comments will apply equally to many other underlying variables.) In order to measure attitude$_1$, we specify certain items, and, by using scale analysis on the responses, obtain a measure of attitude. At this point, we have explicitly or implicitly, introduced a new definition of the underlying variable, say attitude$_2$, as a population of items of which the scale items are a sample. Guttman's "delimited totality of behavior" is essentially the infinite population of items corresponding to attitude$_1$ and would constitute attitude$_2$ if it could be sufficiently specified so that we know the exact items of which it consists. Thurstone's and Likert's initial collections of items are illustrations of specifications of attitude$_2$ but probably consist of too few items to represent attitude$_1$ satisfactorily. There are several reasons why it would seem highly desirable to try to specify attitude$_2$, explicitly and in detail, in measuring attitude$_1$. Among these, are:

(1) we could learn to what extent apparent differences in attitude in different studies are due to differences in attitude$_2$ or in the sample of items from attitude$_2$.

(2) we could test to see whether different specifications of attitude$_2$ are equivalent.

(3) we could use sampling techniques for the selection of items for the scale, instead of relying on largely subjective methods.

Thurstone's method of "equal appearing intervals" essentially consisted in a specification of attitude$_2$, then a stratification of the items of attitude$_2$, and, finally, a technique of sampling the items. Other sampling techniques of a more objective kind could be used, so that the construction of equivalent

scales and the reduction of the error of measurement could be achieved. From these samples of items we could, by using the Guttman and Lazarsfeld scaling techniques, obtain inferences concerning attitude$_2$. Then the only question would be the extent to which attitude$_2$ represented attitude$_1$. It is important to realize that we do not eliminate this question by ignoring it. Even if the existence of a scale is indicated by any sample of items from a scale, it may be that the selected items do not adequately represent the entire range of attitude$_1$ or any reasonable attitude$_2$. Even if different specifications of attitude$_2$ are appropriate in different cultures and at different times, it would be better to make these changes in attitude$_2$ and then sample the items than to pass over the changes in the hope and expectation that the existing sample of items is close enough.

It should be noted that for the above definitions of attitude$_1$ and attitude$_2$ both Guttman and Lazarsfeld deal with attitude$_1$ but not attitude$_2$. The selection of items from attitude$_1$ is not comparable to that of a random sample from an infinite population, since the selection is not objective and does not involve known probabilities of obtaining any particular sample.

It is sometimes said that the sampling of items is difficult. The reason for this difficulty lies, however, not in the lack of appropriate sampling methods, but in the lack of definition of the population of items. To define this population as attitude$_2$ may take some time for any particular attitude, but, once done, the selection of a sample of items will not be particularly difficult. If it cannot be done, then I doubt that attitude$_1$ itself has been sufficiently defined to serve as a basis for selecting items in the manner now used.

Thus, the underlying variable, attitude$_1$, is essentially the unknown population of which a sample of items has been selected by the scale. Inasmuch as we can use in sample design any knowledge that is available concerning the population, we should, through the objective sampling of items, be able to improve on the methods now in use.

The studies of prediction

Chapters 13 through 16 deal with two case studies in prediction. Chapters 13 and 14, prepared by Shirley A. Star, deal with the screening of psychoneurotics in the army and Chapters 15 and 16, prepared by John A. Clausen, deal with the post-war plans of soldiers. In the first problem of prediction, that of screening psychoneurotics, a pencil and paper test was to be used to reduce the number of persons that psychiatrists would need to examine, or, at least, divide the entire group of men receiving pre-induction physical examinations into those who would need cursory psychiatric examination and those who would need intensive psychiatric examination. One result of this study is that it makes clear that the performance of psychiatrists was much more variable than the performance of the test. It is likely that, due to the conditions under which the induction centers operated, it was impossible for independent psychiatric checks to be made of the work of the psychiatrists, but from the data presented, it is almost certain that any such

check would have shown that the variance between the judgments of two psychiatrists on the same man would have exceeded the average variance between the psychiatrist and the psychoneurotic test. Thus, it would have been preferable not to call the judgments of psychiatrists the "criterion" variable. Furthermore, if the objective of the examination was both to keep psychoneurotics out of the army and to prevent injustice to the men who were either included or excluded, then again the data presented concerning the examinations by psychiatrists almost certainly indicate that the objective would have been at least as well attained by the use of the test as by the use of the psychiatrist. It seems unfortunate that the data do not indicate the results of examinations by psychiatrists concerning whether the men would be psychoneurotic for both civilian and military life, or whether they would have been so only for military life. This screening examination could probably have been used not only to reduce the amount of work of the psychiatrists but even to reduce the use of psychiatrists solely to that of an appeals board. The quality control aspects of statistical tests of this sort are, perhaps, not sufficiently realized. When an objective statistical test of the screening type is given, one is essentially performing a relatively cheap inspection of product. It would seem reasonable to use the more intensive examination only in special cases and as a check on the superficial examination, especially when, as in this case, the superficial examination seems to have a more desirable nature than the general examination. Perhaps what could be done in these cases is to treat the screening test as the first test of a double sampling design, and then to sample at different rates from each of several strata for check examinations by psychiatrists, one stratum being entirely examined.

A good many of the statistical points that are made in the two chapters would seem to require more justification than they are given. For example, page 494, no evidence is given to justify the more conservative level of significance as being sufficient to take care of the overestimation of differences resulting from the basis for the dichotomization used. The first principal component is known to have fewer errors of measurement than the multiple regression. However, it has not been shown at what point one should stop, i.e. whether one should use only the first component or the first several components. Furthermore, when Thurstone's centroid method is used, it does not follow that the errors of measurement are minimized in the first factor. In fact, it can be shown that the first principal component is required for the minimization of errors of measurement. Some of the studies, based as they are on the judgments of the psychiatrists and the large variability that seems implicit in the work of the psychiatrists, become dubious when small differences occur. Three major statistical questions occur to one in connection with the screening of psychoneurotics. (1) Does the criterion variable have such large variability of measurement that the statistical measurement is in some sense more satisfactory? (2) How, from a large set of items, should one choose a subset having maximum predictive power? (3) At what point should the screening cut-off take place and what action

of a more intensive nature should occur in order to be certain that losses are minimized and the quality of product not changed? While statistical answers can be given for the first and third questions, the second cannot yet be answered.

Chapters 15 and 16 deal with "Studies of the Post-War Plans of Soldiers," particularly with respect to their occupational plans and intentions. Except for the large amount of resources that were devoted to this problem because of its importance, this is much the same as the problem of deciding whether to buy a certain product. The soldiers were a group of men who had certain hopes and desires which they wished to exercise in the future and who were subject to many obstacles of different kinds in attaining their objectives. It is a useful study to have made and one may hope that follow-up studies of this or similar problems will occur. Until such studies do occur, it is doubtful whether the experience embodied in these chapters should be taken as a basis for generalization. Section 5 of Chapter 16, pp. 675–695, provides a very useful summary of the conclusions reached by Mr. Clausen, but until evidence on the relative costs and variances resulting from the different types of questions and method of prediction are available, it is doubtful whether these conclusions provide useful guides for future studies. Perhaps it may be possible, through studies of this type, at least to attain evidence in the study itself, concerning whether intentions or attitudes are likely to determine behavior. Certainly, measures of intensity might be helpful, but so, also, might be the evidence of consideration of problems as compared with the problems that one may actually face. If, in this type of study, one could ask whether the respondent has enough capital, previous knowledge of business, realization of the possibility of price increases, or unavailability of material, and if one could find out the type of business, the training for that business, and the competition in that business, one could, perhaps, be in a better position to evaluate how likely intention would be to become behavior. Of course, such detail could occur only in a follow-up survey, but in some cases the importance of the results may justify getting this information. There seem to be no important statistical questions to raise concerning this survey, except, perhaps, to wonder why measures of variability were not included.

Conclusion

It is impossible to do justice in a review to the many contributions in this in this volume. Here are given mathematical models that are useful abstractions from the world of experience. Mathematical implications are worked out and applications given. The uses of the techniques in predicting behavior are discussed and illustrated. It is often possible to disagree with some of the details, but not with the conclusion that new, exciting, and important ideas have been united with great insight into a highly stimulating and valuable result.

ERRATUM: ON THE DETERMINATION OF SAMPLE SIZES IN DESIGNING EXPERIMENTS

A. M. Mood
The Rand Corporation

A PAPER with the above title published in this journal (Vol. 43, 1948, pp. 391–402) is based on an erroneous argument The argument is contained in the appendix to that paper which was written by myself. The error is in failing to take account of the fact that n (pp. 400, 401) is a random variable. The result is that the method described in the paper provides merely an approximate method for determining second sample sizes, not an exact method as is implied in the paper.

I am indebted to several of my friends for kindly bringing the error home to me by asking for further elucidation of the argument.

REPRINTS OF ABSTRACTS IN STATISTICAL METHODOLOGY

Edited by MAX A. WOODBURY, *Princeton University*

POPULATION INDEX—1951

1206 U. S. Bureau of the Census.
State economic areas. A description of the procedure used in making a functional grouping of the counties of the United States. By Donald J. Bogue. Washington, 1951. 96 pp.
The text includes a general section on definitions, comparability, etc.; a map of the economic areas and the basis for the system of economic area classification; and procedures for delimiting the state economic areas. Table A, "Nonagricultural characteristics," presents 88 statistical measures for each area. Table B presents basic agricultural statistics for each area. Table C is a list of counties in each state economic area.

1592 Vlach, V.
Application of sampling method by Czechoslovak national census. (Použití výběrové metody pro čs. národní sčítání.) Statistický Obzor 30:44-52. 1950.*

1598 Giaccardi, Fernando.
A critique on the construction of indices of concentration. (Un criterio per la costruzione di indici di concentrazione.) Rivista Italiana di Economia, Demografia e Statistica 4(3-4):527-538. July-Dec., 1950.

1599 Harris, T. E., et al.
Timing of the distribution of events between observations: A contribution to the theory of follow-up studies. Human Biology 22(4):249-270. Dec., 1950.

1602 Winkler, Wilhelm, et al.
Sampling methods in relation to population censuses. (Das Stichprobenverfahren im Dienste von Volkszählungen.) Statistische Vierteljahresschrift 3(3-4):111-122. 1950.
Three brief methodological notes: "Einleitende Bemerkungen," by Wilhelm Winkler; "Stichprobenverfahren bei Bevölkerungserhebungen," by Heinrich Wagner; "Zur bayerischen Volks- und Berufszählung," by Adolf Adam.

PSYCHOLOGICAL ABSTRACTS—1951

2752. Iyer, P. V. Krishna. (*Indian Council of Agricultural Research, New Delhi, India.*) **Runs up and down a lattice.** *Nature, Lond.*, 1950, 166, 276.—A method is briefly presented for calculating the factorial and product moments of distributions arising in the theory of runs up and down on a lattice.—*A. C. Hoffman.*

2745. Bonnardel, R. (*U. Paris, France.*) **Méthode rapide pour le calcul des corrélations moyennes.** (A quick method for calculating average correlations.) *Travail hum.*, 1950, 13, 274–282.—Using conventional methods, average correlations can be obtained only after computing the first-order coefficients, a lengthy procedure. The author presents here a method for computing average correlations directly, with a minimum number of operations and time expenditure. The formula is given, and a sample problem worked out. English summary.—*R. W. Husband.*

3527. Aitken, A. C. (*Mathematical Institute, 16 Chambers Street, Edinburgh.*) **On the Wishart distribution in statistics.** *Biometrika*, 1949, 36, 59–62.—Wishart's distribution is the probability distribution of the estimates of $\frac{1}{2}k\ (k+1)$ moments of the second order, usually called variances and covariances, for a sample of n k-ary vectors drawn from a k-variate normal correlated population. Various methods have been given for the derivation of this valuable distribution, which evidently generalizes to the case of a symmetric and positive definite matrix variate the familiar gamma distribution of a scalar variate. Wishart's distribution is modified; and the moment-generating function of the estimates, matrix transformations and related vector transformations are discussed.—*G. C. Carter.*

3530. Carter, A. H. (*King's Coll., Cambridge U., England.*) **The estimation and comparison of residual regressions where there are two or more related sets of observations.** *Biometrika*, 1949, 36, 26–46.—A technique is developed for estimating and comparing residual regressions where there are two or more related sets of observations. Specific formulae, in convenient form for calculation, are derived for the general case of p correlated samples with q independent variates. The special cases of $p = 2$ and of $q = 1$ are discussed, and the known results for independent samples are briefly deduced and compared. The underlying assumptions are discussed and a comparison of the method with that proposed by Yates is made.—*G. C. Carter.*

3538. Godwin, H. J. (*University Coll. Swansea, Wales.*) **On the estimation of dispersion by linear systematic statistics.** *Biometrika*, 1949, 36, 92–100.—The efficiency of estimates of the standard deviation of a population which are obtained by ranking the observations of a sample and taking a linear combination of them is described. Only the case in which the same rule of combination is used for every sample of a given size is considered. The mean deviation from the mean, which takes different forms according to the position of the sample mean relative to the observations is excluded from the theory. The method used is to consider the first and second moments of differences of consecutive ranks. The most efficient linear systematic statistic and rectangular populations are discussed.—*G. C. Carter.*

ADDITIONAL ABSTRACTS

Guest, P. G., Estimate of the Error at a Point on a Least Squares Curve. *Australian Journal of Scientific Research*, Series *A*, Volume 3, p. 173.

Guest, P. G., Estimation of the Errors of the Least Squares Polynomial Coefficients. *Australian Journal of Scientific Research*, Series *A*, Volume 3, p. 364.

Consider n observations (y_i, x_i) where the model for y's is

$$y_i = \alpha_0 + \alpha_1 x_i + \cdots + \alpha_p x_i^p + w_i$$

with w_i drawn from an infinite population with zero mean, variance $=\sigma_p^2$, and the x_i equally spaced with center m and spacing h. Let $c_{p0}+c_{p1}x+\cdots+c_{pp}x^p$ be the least squares polynomial. The papers give tables facilitating estimation of standard errors of the coefficients c_{pj}.

If r_p^2 is the residual mean square estimate of σ_p^2, then the estimated standard error of c_{pj} may be written

$$r[c_{p,j}] = \frac{\rho_{pj}(k, n) r_p}{h^j n^{j+1/2}} \quad \text{where} \quad k = \frac{2m}{n},$$

$$\rho_{p,j}(k, n) = \sqrt{\sum_{q=j}^{p} \tau_{qj}(k, n)}$$

and the formulae for $\tau_{qj}(k, n)$ are given on p. 175 and p. 366 for $j=0(1)\ q$, $q=0(1)5$.

The values of $\rho_{pj}(k, n)$ are tabulated:

for $|k|<1$ directly to 3 significant figures (Paper 1, Table 1, pp. 176–177; Paper 2, Table 1, pp. 367–371)
for $j=0(1)p$, $p=0(1)5$
$k=0.00(.05)1.00$
$n=7, 8, 9, 11, 14, 18, 25, 40, \infty$
(for some (p, j) more k values, less n values)
for $|k|>1$ in factored form $\rho_{pj}(k, n) \simeq \rho_{pj}(k, \infty) \cdot \phi_p(n)$
$\phi_p(n)$ is tabulated to 3 significant figures (Paper 1, Table 2, p. 178)
for $p=1(1)5$, $n=7(1)16(2)20(5)30(10)50, 100$
$\rho_{p,j}(k, \infty) \equiv \rho_{p,j}(k)$ is tabulated to 3 significant figures (Paper 1, Table 2, p. 178; Paper 2, Table 2, p. 373)
for $j=0(1)\ p-1$, $p=1(1)5$
and $k=1.00(.05)1.20(.1)3.0$ with extra k values for some (p, j).
For $k>3$, $\rho_{p,j}(k)$ is given as $d_{p,j}k^{p-j}$ with $d_{p,j}$ to 3 significant figures.

A discussion of interpolation accuracy is given. The estimated standard error of the jth order differential coefficient with respect to x of the least squares polynomial at the point $x=0$ (m units to left of center of x values) is given by $j!\,r[c_{p,j}]$.

D. L. WALLACE, *Princeton, N. J.*

BOOK REVIEWS

Measurement and Prediction. *Samuel A. Stouffer, Louis Guttman, Edward A. Suchman, Paul F. Lazarsfeld, Shirley A. Star,* and *John A. Clausen.* Volume IV of Studies in Social Psychology in World War II. Princeton: Princeton University Press, 1950. Pp. vii, 756. $10.00.
See the article by William G. Madow, pp. 508–514 in this issue.

What Happens During Business Cycles. *Wesley C. Mitchell* (Late Member of the Research Staff, National Bureau of Economic Research, Inc.). New York:National Bureau of Economic Research, Inc., 1951. Pp. xxxi, 390. $5.00.

C. Ashley Wright, *Standard Oil Company (New Jersey)*

This posthumous work, faithfully edited by Professor Arthur F. Burns and members of the National Bureau's Staff, is the culmination of Professor Mitchell's life work. If it fails to attain the high goal which Mitchell set for himself in the days of his youth and toward which he strove, he would have been the first to acknowledge the fact, as he would have been the last to admit it to be the outstanding accomplishment it is. Few men have set their aim so high, few have planned their work so well, few have persevered so long, and none has followed more meticulously the dictates of a scientific conscience or been more sensitively aware of the traps Truth places in the way of those who follow her. If this volume does not give us an integrated understanding of the cause or causes of the business cycle, which was not its object, it stresses by implication the need for a broader view of business cycle phenomena than is common in much current cycle theory.

The book includes three parts, ten chapters, two appendices, forty-three tables, and seven "charts," two of which contain many figures illustrating the cyclical behavior of particular series, "Chart 1," in fact, extending some eighteen pages. A judiciously balanced introduction by Professor Burns serves to orient the reader and guide him on his way through the detail which follows.

Part I (Chapters 1 to 3, 27 pp.) outlines the task which the National Bureau set itself in the field of business cycle analysis, defines the Bureau's concept of the business cycle, and outlines the methods which it has used in analyzing and collating the mass of data which it has assembled. Characteristically these aims, concepts and methods are attributed impersonally to the Bureau and there is no hint of their relation to Mitchell's life and work. The specialist will find here little not previously available in the monograph "Measuring Business Cycles"[1] beyond the clearest statement the reviewer has seen of the Bureau's use of the clustering tendency of specific-cycle turn-

[1] A. F. Burns and W. C. Mitchell, National Bureau, 1946.

ing dates to determine its peak and trough reference dates (pp. 10–11). The student approaching the Bureau's work in this field for the first time could not do better than begin by reading Burns' introduction and the three chapters of this part before turning to the technical details and more elaborate discussions of the monograph.

Part II (Chapters 4 to 8, 222 pp.) is appropriately named "Varieties of Cyclical Behavior." It discusses in detail variations in timing, conformity, and amplitudes, all between different series, and cycle-to-cycle variability within and among them. It is here that the specialist will find much material which verifies and quantifies more systematically than ever before what has been generally assumed concerning the cyclical behavior of investment, consumers' and producers' purchases, of employment of land, labor, and capital, and of prices and production of farm products, perishable, semi-perishable, semi-durable, and durable goods. The theorist will find little to require a radical revision of his theories, unless they be over-simple in their explanations of cyclical causality, but he will find much reason for correcting and qualifying them in detail. For example, the diversity of inventory behavior (p. 168 and Section E, Table 16) casts much doubt on any assumption of a regular causal role for inventories in the cycle. He will find much that is consistent with his theories, and little that is inconsistent, but much that is consistent will be consistent with other emphases and with conflicting theories. This will be no fault of Mitchell's work but rather because so much theory is stated in forms which are incapable of proof or disproof by data which are available.

The theorist can ill afford to pass this section by or skim it over lightly. It should not be regarded as a mine of information supporting his theories but as a challenge to the nobler and more difficult task of formulating those theories so that they may be tested by its contents.

Perhaps a word of caution is in order. Mitchell's results are, by and large, based on averages and these, as he warns the reader, conceal much significant and wide variation. Mitchell writes for the sophisticated, but even the sophisticated can forget to verify for the particular a general conclusion derived from measures of this sort, especially when in pursuit of an idea which is promising for either its theoretical or practical applications. Illustrative of this is the faith which still exists in selected indicators or the intuitive examination of several indicators as a means of forecasting cyclical turns. Average leads and lags are of little use because the inter-cycle variances of leads and lags are so great for even the best indicators that the latter are of practically no direct help in the prognosis of timing, though it is easy to fall into the trap of believing that they are. Of this sort of danger Mitchell gives us little explicit warning. We must fall back upon that scientific caution with which he was endowed, which he assumes in his readers, and which all of us should cultivate, perhaps less assiduously than he did, as a counterweight to the virtues of bold speculation.

If Mitchell gives insufficient warning of this danger to the unwary, the whole of the section is a warning of a more important sort, namely, against the oversimplification of theories of cyclical causality. The section is devoted to and stresses the almost infinite variety of cyclical phenomena, and more variety is concealed behind the averages it uses. In a sense, perhaps, variety is overstressed for theory must be based on abstractions simple enough to be comprehended, analytically handled, and fruitfully applied. However, in the present state of business cycle theory, this emphasis might well become a crucial antidote.

If Part II emphasizes the varieties of cyclical behavior, Part III (Chapters 9 and 10, 60 pp.) attempts to sort out common features and to combine them into a general description of a typical business cycle, or, more exactly, into a summary of the typical behavior of the individual series in a selected sample as the cycle moves from stage to stage. As Mitchell points out (p. 253), "In large part, this summary repeats what has been known to careful observers, merely putting familiar impressions into definite form (and) shoving under them a more solid foundation. . . . " This statement reflects Mitchell's usual accuracy, but, as he also suggests, the careful reader will find many points of detail which will enable him to refine his views on the timing, amplitudes, accelerations, and decelerations of processes which are presumably important to a causal understanding of the cycle. Many of these details, and those of the preceding part, seem to hint of things to come, of changes which will be necessary in our theories, and of causal relations between at least many of those endogenous variables which are usually assumed to be independent of one another.

This is not the place to revive the ghosts of old methodological controversies. It would smack of presumption and would be premature to attempt an evaluation of Mitchell's contribution to knowledge. But it would be unwarranted and unfair to leave the reader with the impression that his methods and his labors were limited to providing us with qualifications and refinements of what we already know, at least in broad outline.

From the beginning his aim was the critical evaluation and the inductive testing of alternative theories of the cycle. For this end he assembled a mass of material as suitable to his purpose as is possible in a field in which controlled experiments cannot be conducted. Faced with a tremendous task of analysis, he devised techniques which are open to criticism on grounds of statistical precision and "efficiency" but which made his program economically feasible. He tested his methods meticulously, and one lays down this last of his works with the conviction that they were more than equal to the strains he put upon them.

He used alternative theories as a general guide to his thinking but avoided reliance on any single theory. It seems clear that even now no single "literary" theory of the cycle would alone be a reliable guide to research. Mathematical theories are, by and large, the symbolic representations of "literary," theories. They have their own virtues of precision and logical consistency

but they impose limitations of their own through the necessity of mathematical simplification. Because of their dependence upon "literary" theories, the reliability of both for research and policy making is still an open question. Mitchell devoted his life to seeking an answer.

We do not know what his answer would have been. He gave his last years to the search until, realizing that little time was left to him, he turned aside to write the volume under review, lest what he had accomplished might be lost. The task of comparing and revising his results in the light of theory and of testing and revising theories in the light of his results remains uncompleted. Perhaps, it can never be completed in a final sense, but the carefully constructed material of this volume is a challenge to make the attempt. Only when this has been done will it be possible to throw into proper perspective the contribution which this volume, as a part of the broader program Wesley Mitchell set for himself, makes to the advancement of knowledge.

Production Forecasting, Planning, and Control. *E. H. MacNiece* (Director of Quality Control, Johnson & Johnson, New Brunswick, N. J.). New York: John Wiley & Sons, Inc., 1951. Pp. xiii, 305. $5.50.

LORING G. MITTEN, *Ohio State University*

THIS book ranks high on the list of recent publications in the field of production control. The techniques and concepts undoubtedly represent today's better industrial practice and thinking. The book will be of considerable interest and value to production control practitioners. Nevertheless, this reviewer was most disappointed to see another "cook book" appear in a field where practical experience and rules-of-thumb substitute so freely (and unnecessarily) for mathematically formulated analysis. Production control is at present in a state of underdevelopment similar to, or even worse than, quality control was before the contributions of Shewhart. Included between the covers of Mr. MacNiece's book are enough unsolved and badly solved industrial problems to provide a small army of mathematicians and statisticians with material for ten years of research.

The basic concept of production control presented by the author is quite sound. In essence he states that there are a number of pre-production and post-production activities which must be undertaken if costs are to be kept low and delivery promises are to be met. The following sample of the book's twenty-three chapter headings indicates the type of activities which are considered to make up the task of production control: Manufacturing Planning, Factory Planning, The Basis of Production Planning, Sales Forecasting, Procurement, Stores and Raw-Materials Inventory Control, Loading, Scheduling Production, Recording and Reporting, and Evaluating Results and Controlling. A variety of other topics are covered, including the relation of production control to accounting and quality control and also the problems of production control in a war economy. In addition to problems at

the end of each chapter, an extensive practical problem and its solution are presented toward the end of the book. The many flow charts are excellent, but the bibliography is rather inadequate.

The weakness of the book, and of the whole field of production control, lies in the techniques recommended for carrying out the prescribed activities. The chapter on sales forcasting is a good example. The fact that ten pages are devoted to the topic "Techniques of Sales Forecasting" will give some idea of the adequacy of the treatment of this exceedingly difficult problem. By no stretch of the imagination could the adjective "statistical" be applied to any of the extremely crude methods described (although the author titles two of the sub-topics "Statistical Sampling" and "Statistical Projection"). Through an incorrect use of the moving average the author concludes (p. 82) that "these figures show that (the) decline began in 1944" whereas it actually began in 1940 or 1941. The author's "correlation analysis" consists of determining whether an independent and a dependent variable are approximately equal to each other each year over a period of years. If so, the variables are considered to be correlated (leaving the uninformed reader with the impression that the variables are not correlated if they are not approximately equal).

The author's rather uncritical attitude toward the techniques of the field is illustrated by his assertion (p. 32) that the methods-time measurement (M.T.M.) procedure is a "*scientific* method . . . that permits *accurate* time values to be set without the use of a stop watch" (italics are the reviewer's). Actually, the accuracy of the method has not yet been properly investigated. At another point in the book (p. 100) the author states, concerning break-even charts, that "the percentage of profit before taxes increases geometrically with increased volume (above the break-even point)." Actually, the first derivative of the percentage profit with respect to volume is inversely proportional to the square of the volume.

All things considered, the reviewer feels that Mr. MacNiece has done a more than adequate job, relative to the standards of the field. If this volume is of interest to statisticians, however, it will be because of the virgin field it exposes.

World Resources and Industries, A Functional Appraisal of the Availability of Agricultural and Industrial Materials. Revised Edition. *Erich W. Zimmermann.* New York: Harper and Brothers, 1951. Pp. xvi, 832. $7.50.

Most statisticians interested in international economic comparisons and economic development, as well as those concerned only with single industries or resources, will profit from an examination (or re-examination) of this remarkable volume. Zimmermann's emphasis on the social factors that determine resource position and possibilities seems to be even more necessary today than in 1933, when the first edition appeared. The wealth of up-to-date charts and tables in the new edition makes it a handy reference

book; and one can spend a pleasant evening sampling the pictures, interesting facts, and wise observations in this well-printed book.

S. F.

United Nations, Statistical Yearbook, 1949–50, Second Issue. New York: Statistical Office of the United Nations, 1950. Pp. 553. $6.00.

PRACTICALLY all the tables which appeared in *Statistical Yearbook 1948* have been maintained, and quite a few new tables added. Further, many of the existing tables have been considerably expanded to give increased country coverage. However, while the second issue adds data for 1948 and 1949, it drops data for 1928 and 1929. The first issue is therefore by no means obsolete.

S. F.

Elementary Principles of Statistics. *A. C. Rosander.* New York: D. Van Nostrand Co., Inc. 1951. Pp. x, 693. $6.00.

HARRY P. HARTKEMEIER, *University of Missouri*

THIS is an excellent textbook. It has a fresh new tone and the author is quite bold. For example, in the preface he states, "Terms of doubtful value, such as 'kurtosis,' 'probable error,' and 'coefficient of alienation,' have been omitted because they are misleading, superfluous, or not significant in practice." Perhaps this boldness arises from the fact that the author works in the Bureau of Internal Revenue!

The 664 pages of text material consist of 36 chapters divided into four parts: Basic Concepts, Distribution of Measurements, Distribution of Estimates, and Distribution of Test Statistics. There is no treatment of the construction of questionnaires, graphic presentation, statistical tabulation, index numbers, or business periodicities, because the author thinks that the first three belong in a book on the design of statistical inquiry and the last two in a book dealing with the application of statistical principles to economic problems. "This volume is intended to be an introduction to the science of statistics. Although mathematics is, and must be, employed, this is not a book on mathematical statistics." It is a rather long introduction, and I doubt if many instructors will try to cover it all in a one-semester course.

The preface contains a very clear statement of the wide possibilities of the applications of statistical methods in many fields:

> Applications have been drawn from a wide variety of subject matter fields and problem areas in order to show that it is the logical structure of the subject matter or problem, rather than the subject matter itself, which is of basic importance. The principles of statistics have applicability wherever these logical patterns are met; they are not peculiar to economics, or biology, or physics, or psychology, or to any other special field or problem.

On page 6 the author states, "It becomes more and more apparent as statis-

tical theory and statistical practice develop that statistics is a science in its own right." These parts of this book would be especially appropriate as required reading for college and university presidents, deans, and faculty members charged with the responsibility of keeping their institutions abreast of developments in this part of the field of science.

One unusual feature of this book is the author's belief that the science of statistics can be written in terms of twelve basic concepts. These are: population, attribute, variate, sample, frequency distribution, probability, randomness, error, correlation, estimate, hypothesis, inference. The author evidently intended to develop the entire subject around these twelve basic concepts. For a while it is easy to follow this type of organization, but as the more complicated techniques are introduced in the latter part of the book it becomes difficult to keep this organization in mind and to see which one of the basic concepts is the primary one the author has in mind. However, this is not a serious defect. Perhaps the author did not mean to concentrate on just one concept at a time, and regards some techniques as overlapping many concepts.

The author introduces the frequency distribution early, not as a means of organizing or condensing data, but as a basis of making probability statements about estimates and inferences. Although the author states that he has omitted the term "kurtosis," and it does not appear in the index, β_2 and g_2 are introduced as measures of "peakedness." The book includes a treatment of stratified random sampling, biased error, restricted sampling, serial sampling, sampling from a rectangular distribution, and the design of samples. He states, "We are beginning to see that an earlier notion that the planning and collection of statistical data are merely questions of common sense, or involve nonmathematical problems which statisticians can ignore, was a grave mistake."

The book includes a discussion of the binomial, t, chi-square, and z distributions, the analysis of variance with one or more bases of classification, the analysis of covariance and correlation, tests of rank order, and sequential sampling. There are the usual tables, plus several that are not quite so usual, such as Random Numbers and Non-repeating Random Numbers 001–200, 001–500, 001–1000.

Laboratory Manual for Elementary Statistical Methods as Applied to Business and Economic Data. *William A. Neiswanger, Floyd B. Haworth* and *William L. Leavitt.* New York: The Macmillan Company. Pp. iii, 96. $2.50.

ERNEST KURNOW, *New York University*

Too often workbooks and manuals in elementary statistics have been merely "recipe books" for calculating the various descriptive statistical values, giving very little attention to the stimulation of independent thinking by the student. In the case of the present manual, however, the authors are indeed justified in their claim that the outstanding feature of the manual is "the emphasis on analysis and interpretation of results."

Some significant changes have been made in the present revision of the manual, originally published in 1942. Greater emphasis has been given to problems in sampling. The section on simple random sampling and the concept of error has been improved and a special section on the interpretation of sampling statistics has been added. The total number of problems has been increased from 70 to 90, thus making for greater flexibility in the instructor's choice of problems. Some of the problems have the calculations partially completed as an aid to students who lack laboratory facilities.

An appendix has been added containing exercises for beginners with computing machines, tables of areas under the normal curve, square roots, logarithms and random numbers, and a set of simple instructions for the use of the slide rule. These additions should facilitate the work in the laboratory. As heretofore, both arithmetic scale and semi-logarithmic graph paper are provided.

The reviewer doubts whether the one library exercise based on data from the *Survey of Current Business* is sufficient to enable the student to "know the leading sources of data." Furthermore, the choice of titles for a few of the problems in the time series section seems questionable. The fourth problem is entitled "straight-line arithmetic trend." The use of this title for the first time at this point is rather confusing since the student has already been concerned with arithmetic straight line trends in the preceding three problems. In problem 71 the term "geometric" seems to have been inadvertently omitted from the title.

For the instructor who likes to use a manual, the present revision provides a variety of interesting and meaningful exercises to aid students in developing an understanding of the quantitative language. Although the manual was prepared for use with Neiswanger's textbook, it can readily be used with any accepted text in elementary statistics.

Biometric Analysis: An Introduction. *Alan E. Treloar* (Professor, Division of Biostatistics, School of Public Health, University of Minnesota). Minneapolis: Burgess Publishing Company, 1951. Pp. v, 251. $4.50.

JEROME CORNFIELD, *National Cancer Institute*

THIS is an instroduction to statistics based upon lectures given by the author in his elementary courses at the University of Minnesota and constitutes "an experimental revision" of his earlier book, *Elements of Statistical Reasoning*. The scope and treatment are largely Pearsonian, a position which the author is at some pains to defend in his preface.

The first eight chapters are concerned with the descriptive aspects of biometry: frequency distributions, typical values, measurement of variation, skewness and kurtosis. In chapter 9 the concepts of errors of random sampling are introduced with brief discussions of statistics and parameters, sampling distributions, confidence limits, and standard errors. Succeeding chapters cover significance of differences, frequency ratios, their application

to vital statistics and the χ^2 criterion. Three chapters follow on bivariate distributions, the correlation coefficient and its sampling error, and linear regression. The volume concludes with chapters on differences between correlated pairs and χ^2 as a test of independence. With some exceptions the development of the subjects covered is logical and the exposition clear. The treatment is nonmathematical, although simple algebraic derivations are given when possible.

Despite the title the volume is apparently not intended as an introduction to statistical methods used in experimental biology and medicine. This presumably accounts for the omission of such standard biometric topics as analysis of variance, experimental design, bioassay, life tables, and assessment of clinical trials. Many statisticians will question, however, whether the opportunity so offered of a broader and more fundamental discussion of statistical methods has been adequately exploited. Thus, regression is considered only from the point of view of the bivariate distribution, a special case, which is likely to bewilder the student when he is eventually confronted with the many situations in which the bivariate formulation is inapplicable.

In his preface Treloar states "the author of this text is firmly convinced that an initial approach to the study of statistical reasoning by way of the small sample fosters the assimilation of procedural technique at the expense of a satisfactory understanding of the logic of statistical analysis," a point of view with which many will agree. It is doubtful, however, if the "large sample" approach favored by Treloar offers any better opportunity of avoiding this difficulty. Thus, chi-square is illustrated with several examples testing whether a set of data could reasonably be regarded as coming from a normal population. While such examples give the necessary skills in the calculation of chi-square they can scarcely be regarded as a serious discussion of how to test the significance of apparent departures from normality. Similarly, in a comparison of the frequency distribution of males and females by brain weight at autopsy, Treloar concludes that the results "leave no doubt on the issue as to whether, on the average, men have more brains than women." This remark is probably intended as a pleasantry, but the thoughtful student will wonder whether a similar result might not have been obtained by comparing autopsy weights of the heart, liver, or lungs and whether anything more than a sex difference in weight has been demonstrated. This point could have been made without considering the calculating procedures involved in the analysis of covariance.

Some post-Pearsonian developments are included but the discussions are brief and in one case, that of confidence limits, misleading if not incorrect. Nevertheless, the subjects that Treloar has covered he has usually treated well. In the main, this reviewer found himself wishing that Treloar's abundant expository skill had been devoted to a treatment whose scope and emphasis were more nearly biometric, and regretting the frequency of examples whose effect was to teach procedures rather than ideas.

Educational Measurement. *Edited by E. F. Lindquist.* Washington, D. C.: American Council on Education, 1951. Pp. xix, 819. $6.00.

LINCOLN E. MOSES, *Columbia University*

THE editor, in his preface to this work, states that "there has long been an urgent need, particularly in the advanced training of measurement workers, for a comprehensive handbook and text book on the theory and technique of educational measurement". Much of the important work that has been done in this field is unpublished, residing in committee reports, records of testing agencies, etc., and the published literature itself is widely scattered in the journals. This book clearly fills the stated need, and is indeed a remarkably comprehensive exposition of the present status of the art and the theory of educational measurement.

Becuse of the magnitude of the undertaking the book was not written by one man. Each chapter is the work of several named collaborators in addition to an author (or sometimes two). In all, there are seventy-one contributors, including most of the leading workers in this field.

The book is formally divided into three parts: Part One, The Functions of Measurement in Education; Part Two, The Construction of Achievement Tests; Part Three, Measurement Theory. The content of the book can perhaps best be conveyed by listing the chapter titles. The chapters in Part One are: The functions of measurement in the facilitation of learning; The functions of measurement in improving instruction; The functions of measurement in counseling; The functions of measurement in educational placement. The chapters in Part Two are: Preliminary considerations in objective test construction; Planning the objective test; Writing the test item; The experimental tryout of test materials; Item selection techniques; Administering and scoring the objective test; Reproducing the test; Performance tests of educational achievement; The essay type of examination. The chapters in Part Three are: The fundamental nature of measurement; Reliability; Validity; Units, scores, and norms; Batteries and profiles. At the end of each chapter appears a bibliography. In all, perhaps three or four hundred different books and papers are so cited.

It is evidence of good editing—and writing—that the three themes, functions of measurement, construction of tests, and theory of measurement, are not treated independently in the three parts of the book. Rather, the thinking in almost any chapter is materially influenced by all three considerations. The book is carefully written; a good deal of attention has obviously been given to precise phraseology. It is not a cookbook; assumptions are stated, and evaluated. There is a considerable amount of philosophical and logical examination of terms and concepts, which occasionally may leave the reader feeling that somehow fewer words would suffice. The users of this book will profit from the fact that it incorporates many lessons from the broad activity in testing programs in World War II.

The statistician will find the book less satisfying than will the measurement worker. There are few derivations, but many formulae, as is appropriate to the book's purpose. The assumptions (and their practical import) underlying the formulae are given. There are some derivations which seem awkward. There are minor notational annoyances, such as a page of formulae where the reader must repeatedly distinguish between an Arabic and a Roman one—in subscript. This reviewer had the feeling that issues of a statistical nature are not always clearly put. Two exemplary instances of putting statistical issues clearly are: Illustration 60, "A classification of scales of measurement (S. S. Stevens)" and Table 15, "Problems and Sample Techniques for the Linear Combination of Measures". The book does not seem particularly sensitive to basic statistical problems which are only partly solved, as for example scaling; again this is doubtless appropriate to the book's purpose.

This volume should prove valuable as a textbook and as a reference. A worker in the field can find a statement of the current status of any major topic, and a guide to the literature bearing on it. This should facilitate critical evaluation of the present structure of theory and practice, and it should facilitate the progress of research in educational measurement.

The Burden of Diseases in the United States. *Alfred E. Cohn* (Member Emeritus, The Rockefeller Institute for Medical Research) and *Claire Lingg* (Secretary, Former Committee on Research, New York Heart Association). New York: Oxford University Press, 1950. Pp. viii, 129+5 charts in separate slip case. $10.00.

MARGARET MERRELL, *Johns Hopkins University*

THIS book is an attempt to interpret time trends in disease by age and cause. It deals mainly with mortality, only one chapter being devoted to morbidity. Unfortunately the authors show only the most superficial acquaintance with the source material or analyses that have been made of it. They state frankly that they had started the study before they became familiar with the special reports on mortality from the National Office of Vital Statistics, yet they repeatedly claim credit for pioneer work in this field. Thus they justify the book by stating "we believe a beginning must be made in scrutinizing critically the vast amount of statistical information that now lies frozen and uninterpreted in the archives and publications of census bureaus" (p.vi). In an introductory historical sketch which covers the period since Graunt's work in 1662, there are a number of references to the authors' studies but there is only a single reference to Quetelet and no mention of Farr, Greenwood, Frost, Lotka, or Dublin, to cite a few of the outstanding people in the procession of those who have unfrozen and interpreted the data during this time.

The observations are presented in graphs which are uniformly artistic but vary in their soundness. Some are informative but others, notably those

in the separate slip case, present continuous data as if they were discrete, compare rates of dubious comparability, and are so elaborate that they defeat their purpose. The text is a semi-popular interpretation of these graphs which the authors call an essay since it is to some extent opinion.

Although discussion of this type may be stimulating and penetrating, in this case it is a disappointment. It is not well organized and is a mixture of of obvious generalities, loose definitions, and faulty interpretations. Thus crude death rate is defined as the rate "without reference to age at death, or its 'cause', or to any other factor" and yet almost in the same breath we have "the crude death rate at that age". (p. 26) Life expectancy is defined as the "number of years a person can count on surviving". (p. 27) Inferences are drawn from absolute numbers when rates are required and even when rates are used there are flat contradictions in statements. For example, in discussing the old age death rate we have the statement repeatedly made on page 26 that this rate does not change and that "the same number of persons in a healthy community will always be dying in each 100,000 of that age", yet on page 35 we have "What these charts demonstrate is the great fall in the incidence of infectious diseases in younger people and the great rise in the death rate among older people."

Illustrations of inexact and incompatible statements appear on almost every page. These faults are particularly unfortunate since the book is addressed to non-statistical people who may be thoroughly misled.

Selection for Parole: A Manual of Parole Prediction. *Lloyd E. Ohlin* (Research Sociologist, Illinois Division of Correction). Russell Sage Foundation, New York, 1951. Pp. 143. $2.00.

DAVID DURAND, *National Bureau of Economic Research*

THIS is a brief and not very technical account of the use of statistical predictive devices by the parole authorities in Illinois. It bears two important messages for sociologists in general and those concerned with parole problems in particular: (1) an effective statement of why predictive devices are desirable; (2) an authoritative statistical record indicating that such devices can be fairly effective.

To appreciate the need for parole prediction, it is necessary to understand the basic purpose of parole. Here there are conflicting opinions. One view regards parole as a form of clemency, possibly to make amends for a severe sentence, possibly as a reward for good behavior. A second view regards parole as a transition during which the prisoner is assisted in readjusting to normal community life; hence the function of the parole board is to select prisoners most likely to readjust successfully. Dr. Ohlin, who takes the latter view, believes that experience tables can be an invaluable aid in making the selection. He rejects, of course, the idea that experience tables should be the sole criterion. Since statistical prediction is highly mechanical, it is often

ill suited to individual applicants, and the parole board should therefore maintain flexibility. Moreover, the experience table may serve broader purposes than mere selection, as Dr. Ohlin implies and the introduction by E. W. Burgess and Thorsten Sellin explicitly states: "Prediction may be just as important or even more important in parole supervision, but as yet it has not been so employed in a systematic way. It is, however, readily apparent that if men have different probabilities of violation, they could be subject to differences in kind and degree of supervision."

The possibility of using statistical methods effectively in the actual selection of applicatns is attested by the record of 4,941 prisoners paroled by the Illinois penitentiary system between 1940 and 1945. Each of these was given a score on the basis of twelve predictive factors, including home status, work record, and personality rating, among others. The record shows considerable correlation between these scores and the incidence of parole violation. For the group as a whole, 28 per cent were violators—including those who infringed the parole agreement as well as those who committed new crimes. But the parolees with the highest scores, 5 to 10, had less than 3 per cent violations; whereas those with the lowest scores, −5 and −6, had 75 per cent.

Much of this book is devoted to describing the predictive methods now used in Illinois—the basis of the scoring system, the selection of predictive factors, and the use of the experience table. This will provide an excellent introductory account for the general reader; moreover, there are technical appendices and a substantial bibliography for those who want to pursue the subject. But the book itself, even including appendices, has insufficient detail to justify its subtitle "A Manual of Parole Prediction." Possibly this is advantageous, since both the author and the Russell Sage Foundation seem to be nearly as much interested in stimulating further research as in describing achievements to date. The author makes no claim that the described methods are optimal in any sense, and he stresses the need for continuing revision and modification in order to keep parole prediction abreast of changes in basic social conditions.

Measuring Social Need: The Method of Standard Scores Applied to the Measurement of Need in Cuyahoga County. *Virginia Kann White.* Cleveland, Ohio: The Press of Western Reserve University, 1951. Pp. 72. Paper.

F. STUART CHAPIN, *University of Minnesota*

THIS study is to be commended as an example of an awakening interest among social workers in the possibilities of systematic and objective social research to provide a more adequate basis for planning and policy making. It is an MS thesis in social administration in the School of Applied Social Sciences.

The first ten pages of introduction are concerned with the purpose of the

study: to develop a device for measuring the relative need for health, and welfare services of 39 areas (consisting of from one to 25 census tracts each) including the city of Cleveland and its environs. From a list of 25 possible factors, eleven were selected as indices of need: average monthly rental, population density, Visiting Nurse Association cases, Tuberculosis new case rate, infant mortality, sub-standard housing, over-crowded housing, child neglect and dependence, juvenile delinquency, private family services, and relief (general relief and public assistance). The elementary statistical method used to convert the raw data into indices and standard scores is painstakingly described in detail. In this context the 1924 editions of the texts of F. C. Mills and H. L. Rietz are cited, but references to more recent works are also given.

Chapter I supplies an excellent statement of the considerations which guided the selection of the original data. Two indices are based on 1940, one on 1947, one on 1948, and 7 on 1949. The effects of such different bases are not discussed. Some useful tabulations and 14 maps are provided.

Chapter II supplies the author's discussion of the meaning of these data, and contains a correlation matrix for the eleven factors plus a composite index (based on combined z-scores). A table of 10 simple regression equations between dependency-neglect rates as the dependent variable and each of the remaining indices as an independent variable are suggested as a basis for prediction (really only estimation in this case, since no time factor in included). The majority of the correlations are of substantial size or high, and 16 are over ± 0.80. It is not apparent that the author recognizes the fact that all of the correlations are ecological correlations rather than correlations between individual traits, that is, the variables are percentages, ratios or means *descriptive of the properties of groups of individuals* (in census tracts and homogeneous larger areas), and are not descriptive properties of individuals. W. S. Robinson ("Ecological Correlations and the Behavior of Individuals", *American Sociological Review*, 15:3:351–347; June, 1950) has demonstrated that ecological correlations tend to be much higher than the individual correlations which underlie them. In one example he found an ecological correlation of 0.946 between the per cent population 10 years and over which was Negro and the percent of the same population which was illiterate, but the individual correlation on the same data was only 0.203.

Since the author seems unaware of Robinson's critique of ecological correlations, one wonders whether her correlations are dependable and useful, or would fade away when reduced to individual correlations. Although the author discusses "need" in her introduction and is to be commended for limiting the concept to relative need, it is not clear just what the eleven indices measure in need. It would seem to this reviewer that the "need" represented in these indices has three components: (1) need may be regarded as a departure from the average condition or from some more ideal norm, a distinction not clearly made by the author; (2) need as described in this study

may be only response to services, rather than descriptive of the number of persons in want of services; if so, then such relatively over-serviced programs as those of polio and tuberculosis, as compared with relatively under-serviced programs of cancer and cardiac conditions, would weight the results; and (3) since "need" is always and primarily that of individuals, not of geographic groups of individuals or categories, which are matters of convenience for classification, it should be noted that Robinson has shown that the relationships measured in ecological correlations do not adequately portray the individual correlations which underlie them. For this reason the validity of conclusions which accept ecological correlations as a basis for prediction require further analysis and objective test before they may be accepted as a basis for planning and policy making. These comments all apply with equal force to Chapter III, summary and conclusions.

The monograph provides a bibliography of 35 titles: 16 in social work research, 9 in social work and special studies, 7 in statistical method, and 3 in demography. The tabulated data constitute the most useful contribution to research of this study, since the statistical interpretation is elementary and not critical.

Logical Foundations of Probability. ***Rudolph Carnap*** **(Professor of Philosophy, University of Chicago). Chicago: The University of Chicago Press, 1950. Pp. xvii, 607. Originally $12.50, now $8.50.**

BRUNO DE FINETTI, *University of Trieste*

THE way in which Carnap proposes to build a logical notion of probability should be of great interest to everyone concerned with scientific method in general and with statistical method in particular. Some of the clarifying distinctions which he introduces should result in more meaningful discussion among the supporters of the various competing theories of probability. It is not likely, however, that many readers of this Journal will wish to go into the technical details of the logical and semantic constructions involved. It is therefore appropriate that this review confine itself, for the most part, to the general outline of the ideas expressed in the book.

The author is led by his analysis to introduce two different notions of probability, which he calls "probability$_1$" and "probability$_2$." Probability$_1$ means probability in the sense of "degree of confirmation," and probability$_2$ in the sense of "relative frequency." These two notions together may be considered adequate *explicata* of two different meanings of the *explicandum* "probability" in the process he calls *explication* of prescientific concepts. Probability$_1$ is a logical analytical notion, on which the whole of inductive logic can apparently be built; probability$_2$ is a factual notion not by itself capable of supporting a systematic and consistent reconstruction of inductive logic (see, e.g., Sec. 49).

These notions do not altogether reduce the problem of inductive inference to a strictly logical procedure, as the author himself emphasizes: "The demand for a justification of a theory proposed as a rational reconstruction may be understood in two different ways. (1) The first, more modest task is to validate the claim that the new theory is a satisfactory reconstruction of the beliefs in question . . . . (2) The second task is to show the validity of the new theory and thereby of the given beliefs . . . . The system of inductive logic here proposed . . . is intended as a reconstruction restricted to a simple language form, of inductive thinking as customarily applied in everyday life and in science" (p. 576).

The statistical definition (probability$_2$) is subjected to many critical remarks (pp. 343–344, 501, 528, 552–554, 561) concerning difficulties which are usually neglected. To the reviewer it seems that a logical definition (probability$_1$) is subject to even greater difficulties, as will be brought out later. For both notions the question of consistency and of the possibility of developing a perfectly logical foundation is considered open. Until these questions are settled, all discussion, however illuminating, must be regarded as tentative and conducted in the hope that the explication will ultimately become a "definition."

"Whether the attempt can and will be successful remains to be seen" (p. 233).

Logical criticism is very penetratingly applied to sharpen the distinction between the two meanings of all secondary notions, such as "expectation" (see p. 529), depending on whether probability$_1$ or probability$_2$ is chosen as the primary notion.

As Carnap's discussion very clearly shows, two steps in the construction of probability theory may be sharply distinguished. The first step leads to the consideration of all functions $c(h, e)$ additive with respect to h as the set of all probability determinations admissible, in a sense, in the context of the given formal language, where $c(h, e)$ is interpreted as the degree of confirmation of the *hypothesis* h with respect to the *evidence* e. The second step is concerned with the possibility of a rational choice of criteria leading to the selection of a narrower set of theories than those established in Step 1, possibly in the selection of a single additive function as an adequate explicatum for probability.

The distinction between these two steps is not peculiar to Carnap, but is essential to the understanding and discussion of every system, as he points out in his review of the determinations of other authors developed at the end of Chapter V (pp. 337–345) and elsewhere. In his own theory, Carnap's explications of the first step are based essentially on considerations of betting behavior, in accordance with the subjective position. Inductive logic, in particular the relevance of every additional experience to the evaluation of a probability, is well founded with respect to all c-functions. As Carnap makes clear, the evaluation $c(h, e) = r$ is not a factual sentence which, owing to the

increased evidence e', is rejected and replaced by the better one $c(h, e') = r'$. On the contrary, each is to be considered as a logically valid statement of the degree of confirmation of h in the light of the evidence at hand in each case (p. 192; see also p. 299 and p. 555). Inductive reasoning appears here, just as in classical theory, as an application of Bayes' rule (see Sec. 60) which does not require either factual or metaphysical principles such as "uniformity of the world" (pp. 178–182).

The second step, in Carnap's particular theory, is the definition of a narrower set of c-functions, that is, the set of *symmetrical* c-functions (Chap. VIII), in terms of the symmetry properties of sentences in the given language system. Finally, a single function among these, c^* is suggested as a unique determination for probability, in the appendix, where Carnap briefly outlines the problem of a unique determination, which is to be the subject of a second volume. As the author says, "A criticism of the system based on c^* would hardly be useful at the present time, that is, before the publication of the full explanation of this system" (p. 563).

The fundamental question will, however, be this: Does the c^* give a uniquely defined solution for every problem, and if so, is it universally adequate? Since the choice of the language-system in which a practical problem can be formulated seems to allow a wide degree of arbitrariness, doubt may well arise that "the equality of logical status" is really a firmer principle replacing the principle of indifference and not merely a label which can be attached to any situation where the application of that old principle, or what ever other argument may seem to justify a judgment of equality between several probabilities (see, e.g., p. 494, and some examples) would be resorted to.

Though the importance of Carnap's book depends chiefly on its contribution to the theoretical questions already discussed, there are other particularly noteworthy points, e.g., discussions concerning the notions of relevance (particularly Secs. 69 and 71).

Finally, I should mention that the book is very carefully and thoughtfully arranged, with excellent mnemonic and reference aids, including summaries of chapters and sections, indexes, glossary, bibliography, etc., all of which are very clear and practical.

The Nature and Application of Inductive Logic. *Rudolph Carnap.* Chicago: The University of Chicago Press, 1951. Pp. viii, 161–202, 242–279. $1.25. Paper.

ASIDE from a few words of introduction and a condensed bibliography, this is entirely reprinted from the book just reviewed above. The portions reprinted are the first page of the Preface, and Sections 41, 42, 43, 49, 50, and 51, all of which are included in Chapter IV of *Logical Foundations of Probability*. This selection is intended to give a relatively informal survey of the subject matter of the larger book.

L. J. S.

Theory of Probability. *M. E. Munroe.* New York: McGraw-Hill Company, 1951. Pp. viii, 213. $4.50.

GOPINATH KALLIANPUR, *University of California (Berkeley)*

THIS book is written for students with no more mathematical equipment than one year of calculus and is intended to be a formal summary of the developments of modern probability theory.

The definition of probability adopted in the book is the axiomatic one. The motivation for such a procedure is clearly explained by means of examples. There is something to be desired, however, in the way the author has attempted to build up an intuitive background. It is understandable that a discussion of probability based on equally likely events should occur in an elementary textbook but one is somewhat startled to come across a parallel "definition" in the continuous case which reads: "If a point is chosen at random on a line interval of length A, the probability that it is in a given subinterval of length a is a/A." The author's reasons for going into questions of this sort are that they "serve as important connecting links between the mathematical theory and the physical world." In my opinion, the empirical basis of the axiomatic definition of probability is the frequency interpretation and it is possible to explain the latter without having to introduce the undefined concept of "equal likelihood." This criticism might be dismissed as fastidious, but I believe that a student making his first acquaintance with the subject might well be bewildered by these various "definitions" of probability.

Chapters 3 to 6 deal with such topics as joint distributions, repeated (Bernoulli) trials, convolution of distribution functions, and moments. Properties of Beta and Gamma functions, the derivation of Stirling's approximation to $n!$, and a few other selected results of advanced calculus are gathered together for the convenience of the student in Chapter 7. A discussion of the Riemann-Stieltjes integral could have prepared him for future work with the Lebesgue-Stieltjes integral. A special feature of the book which distinguishes it from most elementary textbooks on probability is the considerable attention given to limit theorems. No less than three chapters (8, 9, and 11) are devoted to an informal treatment of the laws of large numbers and the classical central limit theorem. The chapter on the latter is perhaps the most interesting in the whole book. It contains the statement of the Lindeberg-Feller theorem from which various special forms are deduced as corollaries. An actual proof of the central limit theorem is given in the case of independent and identically distributed random variables. Bernoulli trials, in particular, are discussed in great detail. Chapter 8 provides a general introduction to limit theorems. There is even a brief reference to the law of the iterated logarithm. The mathematical setting for a treatment of the strong law, viz. the space of infinite sequences of trials is introduced with reference to a particular example (coin tossing) and the relation between the strong and the weak laws explained.

The following are some of the simple applications to sampling theory considered: the convergence in probability of the sample mean and variance to the corresponding values in the population; the asymptotic normal distribution of the sample mean and a brief discussion of the chi-square test. Numerous examples are provided at the end of each chapter and in almost every case the answers are given, often with hints for solution.

A minor point worth mentioning, and one which might cause confusion in the mind of the student, is the unfortunate use of the same letter in two different senses, to denote a random variable as well as an ordinary real variable occurring as the argument of a function. Indeed this usage has resulted in such a statement as the following: "We have defined a function $f(x)$ of the stochastic variable x . . . " (p. 21).

The mathematics is mostly of a formal nature and as the author takes every opportunity to emphasize this fact, it is perhaps unfair to criticise any shortcomings of the book in this direction. Nevertheless, one feels that the definition of probability in the continuous case could have been more satisfactorily treated. This is admittedly a very difficult task without the use of measure theoretic notions. In Chapter 2, for instance, the student who has an acquaintance with the ordinary Riemann integral will puzzle over the symbol $\int_E f(x)\,dx$, which is introduced without an explanation.

On the whole, the author's object has been to enable the student to gain a quick familiarity with the important results of the theory of probability. The average student will find the book useful and interesting.

Trabajos de Estadistica, Vol. I, 1950. Madrid.

THIS is the first volume of a new Journal published by the Departamento de Estadistica of the Consejo Superior de Investigaciones Cientificas. This review plans to include articles on pure or applied statistics. This issue contains articles by F. Azorin and H. Wold; E. Cansado, E. Cansado and H. Wold, A. Guiraum, J. Tena and H. Wold, and H. Wold. It also has sections for notes, information, and bibliography. The price is three dollars annually.

M. A. L.

Revue Statistique, Numero 1, Mars 1951.

A SPECIAL Yugoslav statistical journal of which this is the first issue is now being published in Belgrade. The articles are in Serbian, with short summaries in French and English. This issue includes five long articles and ten pages of bibliography, a total of 181 pages. The Editorial Board is composed of three members of the Federal Statistical Office, Belgrade, and two professors on the Faculties of Economics at Zagreb and at Belgrade. Articles in this issue are by Stane Krašovec, Dolfe Vogelnik, Miloš Macura, Ivo Lah, and Juraj Ećimović.

M. A. L.

Revue Economique, No. 1, May 1950. Librairie Armand Colin, Paris.

THIS Journal, which is to appear five times a year with a Board of Directors headed by Albert Aftalion of the Faculté de Droit, Paris, is intended to supplement rather than to replace the *Revue d'Économie Politique*, and will be divided into two distinct parts, one on "Problèmes de Méthode" and the other entitled "Chroniques." In this issue, articles in the first part are by André Marchal, Fernand Braudel, Jean Lhomme, Jean Gottman, and Jacques Lecaillon, while Alexandre Chabert, Jean Weiller, Jean Lhomme Alfred Sauvy, and André Piatier have contributed to the second section.

The annual foreign subscription to the *Revue Economique* is 1,600 fr.

M. A. L.

PUBLICATIONS RECEIVED

American Society for Quality Control. **Quality Control Conference Papers, 1951; Fifth Annual Convention.** Paper. $3.50.

Baumol, William J. **Economic Dynamics.** The Macmillan Co., New York. 1951. $5.00.

Clemence, Richard V. **Income Analysis.** Addison-Wesley Press, Cambridge, Mass. 1951. $2.50.

Colombo, Bernardo. **La Recente Inversione nella Tendenza della Natalità.** Casa Editrice Dott. Antonio Milani, Padua. 1951. Paper.

Council of Economic Advisors. **The New England Economy.** Report to the President. Government Printing Office, Washington, D. C. July 1951. Paper. 55 cents.

Dixon, Wilfred J. and Frank J. Massey, Jr. **An Introduction to Statistical Analysis.** McGraw-Hill Book Co., New York. 1951. $4.50.

Dwyer, Paul S. **Linear Computations.** John Wiley and Sons, New York. 1951. $6.00.

Hall, Challis A., Jr. **Effects of Taxation, Executive Compensation and Retirement Plans.** Harvard University, Graduate School of Business Administration. 1951. $4.25.

Hart, Albert G. **Defense without Inflation.** The Twentieth Century Fund. 1951. $2.00.

Herring, John P. **Labor Force, Employment, and Unemployment, Annual Estimates by States, 1900–1940.** University of Washington Press, Seattle. 1951. Paper.

Iowa State College, Agricultural Experiment Station. **Random Arrangements for Lattice Designs.** Special Report #5. Ames, Iowa. November 1950. Paper, spiral bound. $1.00.

Katona, George. **Psychological Analysis of Economic Behavior.** McGraw-Hill Book Co., New York. 1951. $5.00.

Koopmans, Tjalling C., ed. **Activity Analysis of Production and Allocation.** Cowles Commission Monograph #13. John Wiley and Sons, New York. 1951. $4.50.

Leontieff, Wassily W. **The Structure of the American Economy, 1919–39.** Second Edition, enlarged. Oxford University Press, New York. 1951. $5.75.

Linder, A. **Statistische Methoden für Naturwissenschafter, Mediziner und Ingenieure.** Zweite, erweiterte Auflage. Verlag Birkhauser, Basel. 1951.

Lundberg, Erik, ed. **Income and Wealth —Series I.** Bowes and Bowes, Cambridge, England. 1951. 30 s.

Luts, Friedrich and Lloyd W. Mints, eds. **Readings in Monetary Theory.** The Blakiston Company, Philadelphia. 1951. $4.75.

Morse, Philip M. and George E. Kimball. **Methods of Operations Research.** John Wiley and Sons, New York. 1951. $4.00.

Mudgett, Bruce. **Index Numbers.** John Wiley and Sons, New York. 1951. $3.00.

Nutter, G. Warren. **The Extent of Enterprise Monopoly in the United States,1899–1939.** University of Chicago Press, Chicago. 1951. Paper.

Paden, Donald W. and E. F. Lindquist. **Manual for Statistics for Economics and Business.** McGraw-Hill Book Co., New York. 1951. Paper.

Quenouille, M. H. **Introductory Statistics.** Butterworth-Springer, Ltd., London. 1951. 30 s.

Sanders, Thomas H. **Effects of Taxation on Executives.** Harvard University. Graduate School for Business Administration. 1951. $3.25.

Seltzer, Lawrence H., assisted by Selma F. Goldsmith and M. Slade Kendrick. **The Nature and Treatment of Capital Gains and Losses.** National Bureau of Economic Research. New York. 1951. $7.50.

Sibley, Aldrige. **Support for Independent Scholarship and Research.** Social Science Research Council, New York. 1951. Paper. $1.25.

Spiegel, Henry William. **Introduction to Economics.** The Blakiston Company, New York. 1951. $5.00.

Switzerland, Federal Statistical Bureau. **Movement of the Population in Switzerland, 1945–1948.** Two Volumes in French and German. Bern. 1950–51. Paper.

Thomsen, Frederick Lundy. **Agricultural Marketing.** McGraw-Hill Book Co., New York. 1951. $6.00.

Tinbergen, J. Translated by H. Ryken van Olst. **Econometrics.** The Blakiston Company, Philadelphia. 1951. $4.50.

United Nations. **Demographic Yearbook, 1949–50.** Second Issue. Columbia University Press, New York. 1951. $6.00.

U. S. Bureau of the Census. **Catalog of U. S. Census Publications, 1790–1945.**

———. **Census Publication Catalog and Subject Guide, 1950.**

———. **City Government Finances in 1949.**

———. **Foreign Commerce and Navigation, 1946. Volume I, Parts A and B; Volume II.**

———. **Handbook of Statistical Methods for Demographers.**

———. **State Economic Areas.**

———. **State Government Finances in 1950.**

Outstanding McGRAW-HILL Books

ELEMENTS OF STATISTICAL METHOD. New 3rd Edition

By ALBERT E. WAUGH, University of Connecticut. In press

A revision of this outstanding text. While changes have been made in the new edition, the text is still designed to introduce the student to statistical concepts and nomenclature and to encourage him to think in statistical terms. Every effort is made to keep the discussion at the beginner's level and to present basic ideas in such a manner that the student will find it easy to continue under his own power.

SAMPLING INSPECTION BY VARIABLES

By ALBERT H. BOWKER and HENRY P. GOODE, Stanford University. Ready in January

Presents a comprehensive set of sampling plans for use when inspection of the item is by variables and when the lot is to be evaluated in terms of percentage defective. The largest part of the book is practical, giving plans and procedures intended as working material for industrial use, while the rest of the text is theoretical, intended for mathematicians and statisticians interested in sampling inspection theory and practice.

HIGH SPEED COMPUTING DEVICES

By ENGINEERING RESEARCH ASSOCIATES. 451 pages, $6.50

Introduces the reader to the general character of computing machines and the arithmetic techniques employed in using them. Shows the basic circuits used to perform various functions in a computing machine, and the relationship between machine capabilities and the required mathematical tools. Intelligible and up to date, this book will aid the reader in deciding which particular type of computing machine is adaptable to his particular problem.

QUALITY CONTROL HANDBOOK

Edited by J. M. JURAN. *McGraw-Hill Industrial Organization and Management Series.* 800 pages, $10.00

The whole story of quality control presented in a thorough comprehensive form, this book provides—authoritative answers to questions on any phase of planning or carrying out a quality control program; a basic or advanced review of all phases of the quality control function; and quick reference to many formulas, data, record forms and other practical information.

Send for copies on approval

McGRAW-HILL BOOK COMPANY, INC.

330 WEST 42ND STREET, NEW YORK 18, N. Y.

BIOMETRIKA

Volume 38 CONTENTS *Parts 3 and 4*

Biometrika 1901-1951. By W. P. ELDERTON

The Jacobian of certain matrix transformations useful in multivariate analysis. By W. L. DEEMER and I. OLKIN

A chart for the incomplete Beta Function and the cumulative binomial distribution. By H. O. HARTLEY and E. R. FITCH

The effect of standardization on a X^2 approximation in factor analysis. By M. S. BARTLETT

Some systematic experimental designs. By D. R. COX

On estimating the size of mobile populations from recapture data. By N. T. J. BAILEY

The comparison of several groups of observations when the ratios of the population variances are unknown. By G. S. JAMES

On the comparison of several mean values: an alternative approach. By B. L. WELCH

Tables of symmetric functions: Pts. II and III. By F. N. DAVID and M. G. KENDALL

A mathematical theory of animal trapping. By P. A. P. MORAN

Two applications of bivariate k-statistics. By B. M. COOK

Expected frequencies in a sample of an animal population in which the abundances of species are lognormally distributed: Pt. I Theory, Pt. II Application. By P. M. GRUNDY

The fitting of polynomials to equidistant data with missing values. By H. O. HARTLEY

The delay to pedestrians crossing a road. By J. C. TANNER

Interrelations between certain linear systematic statistics of samples from any continuous population. By G. P. SILLITTO

Truncated log-normal distributions: I Solution by moments. By H. R. THOMPSON

Further applications of range to the analysis of variance. By H. A. DAVID

The estimation of population parameters from data obtained by means of the capture-recapture method: I The maximum likelihood equation for estimating the death rate. By P. H. LESLIE and DENNIS CHITTY

Miscellanea: The Wishart distribution by solving simultaneous linear differential equations. By H. S. STEYN. Note on the inversion theorem. By J. GIL-PELAEZ. Note on the variate differences of autoregressive series. By BENTO MURTEIRA. Quality control systems based on inaccurately measured variables. By W. M. LONG. Notes on approximations to the integral of the normal bivariate distribution. By J. H. CADWELL. Mean and variance of an entry in a contingency table. By W. L. STEVENS. Rapid methods for estimating correlation coefficients. By L. N. CHOWN and P. A. P. MORAN. Moment constants for the distribution of range in normal samples. By H. O. HARTLEY and E. S. PEARSON. A note on the statistical independence of quadratic forms in the analysis of variance. By J. A. NELDER.

Reviews: I. J. GOOD'S "Probability and the Weighing of Evidence." N. L. JOHNSON and H. TETLEY's "Statistics" Vol. II.

Please mention the Journal of the AMERICAN STATISTICAL ASSOCIATION *in writing advertisers*

New Irwin Texts for Your Courses in Statistics

QUALITY CONTROL AND INDUSTRIAL STATISTICS

BY ACHESON J. DUNCAN, THE JOHNS HOPKINS UNIVERSITY

This book offers a thorough treatment of those parts of statistical theory which are especially applicable to quality control procedures and industrial research. It is designed to give the student an understanding of statistical principles so that he will not only have a fuller understanding of the quality control procedures which are in current use, but he will also be able to adapt new methods as they arise. In reviewing the fundamentals of probability and statistical theory, the book discusses the meaning of probability, the analysis of frequency distributions, and sampling theory. It treats acceptance sampling by attributes, and by variables; it discusses control charts; it explains the Lot-Plot method and describes binomial probability paper and its uses in quality control; and it covers those statistical methods that are pertinent to the evaluation of test data. It also includes theories of estimation, confidence intervals, tests of hypothesis, analysis of variance, and simple and multiple regression correlation. Military Standard 105A is fully covered. There are over 300 problems in the book and at the end of each chapter is a select list of references.

READY JANUARY 1952

MEASURES OF BUSINESS CHANGE

A Baker Library Index

BY ARTHUR H. COLE, HARVARD UNIVERSITY

This new book encompasses those indicators of economic and business change relating to the country as a whole and those that pertain to specific areas or locations within the country. The first part covers national measures of: business activity, production, and sales; wholesale and retail commodity prices, and cost of living; materials, wages, building and equipment costs; earnings, hours, turnover, and work stoppages; and stock prices and yields, bond prices and yields, and other financial data. Part two is concerned with regional and local indicators relating to the following areas: New England states; Middle Atlantic states; Southern states; North Central states; Northwest Mississippi Valley states; Central Mississippi Valley states; Southwest states; and the Pacific Coast states. A brief description of each index is given, as well as information on the compiler, frequency of publication, location of current and historical data, and the length of period for which the historical data are available. Every effort has been made to cover all indicators prepared by private enterprises and those major indicators which are published by various governmental agencies.

READY JANUARY 1952

RICHARD D. IRWIN, INC.
HOMEWOOD, ILLINOIS

Please mention the Journal of the AMERICAN STATISTICAL ASSOCIATION *in writing advertisers*

ECONOMETRICA

Journal of the Econometric Society

Contents of Vol. 19, October, 1951, include:

OSKAR MORGENSTERN .....Abraham Wald, 1902-1950
ABRAHAM WALD .....On Some Systems of Equations of Mathematical Economics (tranlated by Otto Eckstein)
LAWRENCE R. KLEIN .....Estimating Patterns of Savings Behavior from Sample Survey Data
KENNETH J. ARROW .....Alternative Approaches to the Theory of Choice in Risk-Taking Situations
TJALLING C. KOOPMANS .....Efficient Allocation of Resources
RENÉ ROY .....La Demande des Biens Indirects

List of Members of the Econometric Society

Geographical List of Members and Subscribers

Book Reviews and Notices of Meetings

Published Quarterly ***Subscription to Nonmembers: $9.00 per year***

The Econometric Society is an international society for the advancement of economic theory in its relation to statistics and mathematics.

Subscriptions to *Econometrica* and inquiries about the work of the Society and the procedure in applying for membership should be addressed to William B. Simpson, Secretary, The Econometric Society, The University of Chicago, Chicago 37, Illinois, U.S.A.

BULLETIN
OF THE OXFORD UNIVERSITY
INSTITUTE OF STATISTICS

The BULLETIN publishes articles by members of the research staff and other economists on problems of applied economics and economic statistics.

Some recent articles include:

The Distribution of Capital in Private Hands in 1936-38 and 1946-47 .....K. M. Langley
Family Expenditure in 1949 .....T. Schulz
National Income in the 3rd Quarter .....D. Seers and others
Machinery Prices between the Wars .....H. J. D. Cole
Exchange Rates and National Income .....T. Balogh and P. P. Streeten
Differences between the Wages of Skilled and Unskilled Workers, 1880-1950 .....K. G. J. C. Knowles and D. J. Robertson

Published Monthly *Subscription Price $4.50*

Inquiries and subscriptions to:

THE BULLETIN

Basil Blackwell, 49 Broad Street, Oxford

Please mention the Journal of the AMERICAN STATISTICAL ASSOCIATION in writing advertisers

Princeton UNIVERSITY PRESS

Cohort Fertility
of Native White Women in the United States

By P. K. WHELPTON. To be published, Winter 1952, this study develops new techniques of measuring past fertility trends, and provides an improved basis for forecasting future trends. Special attention is given to the extent to which births are "postponed" during certain times and "made up" or "borrowed from the future" during others. The author, formerly Associate Director of the Scripps Foundation, which sponsored the study, is now Director of the Population Division of the United Nations. *$4.00*

On the Accuracy of Economic Observations

By OSKAR MORGENSTERN. This distinguished economist, co-author of *Theory of Games and Economic Behavior,* presents a study of the principles governing the accuracy of quantitative economic observations. "A stimulating discussion of the disturbing variety and shockingly large size of errors in our basic data."—*Journal of the American Statistical Association. $2.00*

At your bookstore, Princeton University Press

Please mention the Journal of the AMERICAN STATISTICAL ASSOCIATION *in writing advertisers*

POPULATION STUDIES

A JOURNAL OF DEMOGRAPHY

Editor D. V. GLASS

Vol. V, No. 1 CONTENTS **July 1951**

A. SAUVY. Some Lesser Known French Demographers of the Eighteenth Century: De La Morandière, de Caveirac, Cerfvol and Pinto.

L. T. BADENHORST. Population Distribution and Growth in Africa.

J. B. DOUGLAS. Social Class Differences in Health and Survival During the First Two Years of Life; the Results of a National Survey.

COLIN CLARK. Regional Differences in Reproductivity in England and Wales.

F. J. BUCKATZSCH. The Constancy of Local Populations and Migration in England before 1800.

D. V. GLASS A Note on the Under-Registration of Births in Britain in the Nineteenth Century.

Subscription price per volume is $5.00 net, post free

Published for the POPULATION INVESTIGATION COMMITTEE by

CAMBRIDGE UNIVERSITY PRESS
32 East 57th St., New York 22

ESTADISTICA

Official Journal of the
Inter American Statistical Institute

Vol. IX, No. 33 **December 1951**

CONTENTS

Exámen de las Tabulaciones Recomendadas sobre Fecundidad .....
Tabulaciones Especiales sobre Analfabetismo .....
Problemas Relacionados con la Tabulación de Datos sobre Desempleo .....
Apuntes sobre el Plan para la Preparación de Atlas Censales Usando los Datos del Censo de las Américas de 1950 .....Alford Archer
Utilización de los Datos Censales en el Análisis de Mercados ....Philip M. Hauser
Medidas para Ayudar a la Coordinación Nacional de la Estadística .....
A Program for Developing Industrial Statistics .....
Spanish Coding Manual for Regional Application of the Standard International Trade Classification .....
Censos Industriales de las Naciones Americanas—III .....Newton B. Knox
Reports on the I Session of the Committee on Improvement of National Statistics, and on the IV Session of the Committee on the 1950 Census of the Americas, Washington, D.C., June 2-15, 1951 .....
Notas Editoriales. Actividades del Instituto. Noticias Estadísticas. Publicaciones ....

Francisco de Abrisqueta
Editor of the Journal
Inter American Statistical Institute
c/o Pan American Union
Washington 6, D.C., U.S.A.

PROBABILITY AND THE WEIGHING OF EVIDENCE

By I. J. Good

120 pages **Cloth $3.00**

This book aims to provide a simple, consistent, and logical theory of probability to serve as a basis for scientific induction, the use of statistics, and ordinary reasoning.

Other Hafner Books on Statistics:

YULE and KENDALL

An Introduction to the Theory of Statistics

725 pages $7.50

F. YATES

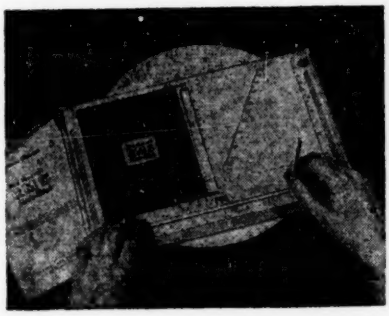
Sampling Methods for Censuses and Surveys

318 pages $5.50

HAFNER PUBLISHING CO., Inc.
31 East 10th Street, New York 3, N.Y.

How to Simplify Your Statistical Work

Cut the time, cost and effort of your statistical computations by using



THE MERRILL RMS SLIDE-DISK

HUNDREDS OF STATISTICIANS are now learning how quickly, easily and accurately their calculations can be performed with this new graphical instrument. Root-mean-squares, standard deviations, geometric & harmonic means, correlation coefficients, etc. are found in a fraction of the usual time and to one-half percent accuracy. Priced at $64 complete. Fully guaranteed.

SEND FOR descriptive brochure and coupon for free RMS Slide-Disk trial.

GRAPHIC CALCULATOR CO. Dept. SA 633 Plymouth Court, Chicago 5, Ill.

1951 OFFICERS AMERICAN STATISTICAL ASSOCIATION

President: Lowell J. Reed; *President-Elect:* Aryness Joy Wickens; *Vice-Presidents:* Dorothy S. Brady, Morris H. Hansen, Philip M. Hauser; *Directors:* Gertrude M. Cox, W. Edwards Deming, Cyril H. Goulden, John W. Tukey, Ralph J. Watkins, Louis L. Thurstone; *Secretary-Treasurer:* Samuel Weiss. *Members of the Council:* C. M. Armstrong, C. I. Bliss, W. S. Brush, Jerome Cornfield, Lester R. Frankel, Paul G. Hoel, J. W. Hopkins, Palmer O. Johnson, H. L. Jones, T. J. Mills, Frederick Mosteller, P. R. Rider, J. R. Stockton, E. J. Swan, W. Allen Wallis, Sylvia Weyl.

PAST PRESIDENTS

1839–45 Richard Fletcher
1846–51 George C. Shattuck
1852–82 Edward Jarvis
1883–96 Francis A. Walker
1897–1909 Carroll D. Wright
1910 S.N.D. North
1911 Frederick L. Hoffman
1912 Walter F. Willcox
1913–14 John Koren
1915 E. Dana Durand
1916 Charles P. Neill
1917 Allyn A. Young
1918 Wesley C. Mitchell
1919 Joseph A. Hill
1920 George E. Roberts
1921 Carroll W. Doten
1922 William S. Rossiter
1923 Warren M. Persons
1924 Louis I. Dublin
1925 R. E. Chaddock
1926 Leonard Ayres
1927 Edmund E. Day
1928 Carl Snyder
1929 Edwin B. Wilson
1930 Malcolm C. Rorty
1931 William F. Ogburn
1932 Irving Fisher
1933 Stuart A. Rice
1934 Frederick C. Mills
1935 Willford I. King
1936 Joseph S. Davis
1937 W. Randolph Burgess
1938 R. H. Coats
1939 Raymond Pearl
1940 F. Leslie Hayford
1941 Winfield W. Riefler
1942 Alfred J. Lotka
1943 E. A. Goldenweiser
1944 Helen M. Walker
1945 Walter A. Shewhart
1946 Isador Lubin
1947 Willard L. Thorp
1948 George W. Snedecor
1949 Simon Kuznets
1950 Samuel S. Wilks

A cumulative Index to Volumes 1–34, 1888–1939, and Annual Indexes thereafter, may be obtained from the Office of the Secretary of the American Statistical Association.

Reprints of all articles published since 1945 may be purchased from the office of the secretary. Articles which have appeared during the past twelve months are usually available in quantity.

Over 95,000 Students Have Been Trained With This Standard Text . . .

APPLIED GENERAL STATISTICS

by Frederick E. Croxton, Columbia University; and Dudley J. Cowden, University of North Carolina

- A remarkably comprehensive treatment of statistical methods and their application to many fields.
- Illustrative material has been drawn mainly from the fields of economics, sociology and business, and are all actual examples, rather than artificial cases contrived for the purpose. Data used for the more than 250 charts are the result of actual investigation.
- Readily understandable and easily teachable. Elementary algebra is the only mathematical requirement.
- *Teaching Aids:* 257 charts illustrate various procedures. 180 tables include those for areas of normal curve, areas of skewed curve, the t distribution, chi square, the z distribution, and F for the analysis of variance, and a new flexible calendar for use in adjusting time series.

944 pages 5½" x 8½" Illustrated

Workbook in Applied General Statistics

Third Edition

by Frederick E. Croxton

Newly revised, this workbook contains 30 exercises for elementary course, plus 8 supplementary exercises which allow for added emphasis on given topics. All exercises are either new, or have been brought completely up to date, and all are *based on actual data*. Clearly shows the development of the various steps and their mutual relationship. Includes all necessary worksheets, graphic and tabular forms, formulas. Although keyed to APPLIED GENERAL STATISTICS, this Workbook can be used with any recognized statistics text.

181 pages text 62 worksheets 8½" x 11"

Answer book free on adoption (Restricted to teachers)

Send for Your Copy Today

PRENTICE-HALL INC. 70 FIFTH AVE · NY 11 · NY

GEORGE BANTA PUBLISHING COMPANY, MENASHA, WISCONSIN